A CROSS-LAYER DESIGN OF WIRELESS AD-HOC NETWORKS

THÈSE N^{O} 3301 (2005)

PRÉSENTÉE À LA FACULTÉ INFORMATIQUE ET COMMUNICATIONS

Institut de systèmes de communication

SECTION DES SYSTÈME DE COMMUNICATION

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Bozidar RADUNOVIC

B.Sc. in Electrical Engineering, Université de Belgrade, Yougoslavie
et de nationalité serbe et monténégrine

acceptée sur proposition du jury:

Prof. J.-Y. Le Boudec, directeur de thèse
Prof. A. Ephremides, rapporteur
Prof. E. Telatar, rapporteur
Dr L. Truong, rapporteur

Lausanne, EPFL
2005

A Cross-Layer Design of Wireless Ad-Hoc Networks

Božidar Radunović

July 12, 2005

Abstract

We consider a cross-layer design of wireless ad-hoc networks. Traditional networking approaches optimize separately each of the three layers: physical layer, medium access and routing. This may lead to largely suboptimal network designs. In this work, we propose a jointly optimal design of the three layers, and we show a significant performance improvement over the conventional approach.

In the first part of this thesis, our goal is to select appropriate performance metrics for the joint optimization problem. To that respect, we analyze several existing rate-maximization performance metrics for wireless ad-hoc networks: maximizing the sum of rates, max-min fairness and proportional fairness.

We first show with several examples that it is not clear if and how max-min fairness can be defined on each one of the examples. We give a formal proof that the max-min fair rate allocation exists on a large class of sets of feasible rates, among which are the feasible-rate sets of all known ad-hoc networking examples. We also give a centralized algorithm to compute the max-min fair allocation whenever it exists.

Next, we compare the three metrics for ad-hoc scenarios in terms of efficiency and fairness. We prove that, similar to wired networking, maximizing the sum of rates leads to gross unfairness and starvation of all but the flows with the best channel conditions. We also prove that, contrary to wired networking, max-min fairness yields all flows having the same rate, thus causing large inefficiencies. These findings offer theoretical explanations to the inefficiency and unfairness phenomena previously observed in the contexts of 802.11 and UWB networks. Finally, we show that proportional fairness achieves a good trade-off between efficiency and fairness and is a good candidate for a rate-based performance metric in wireless ad-hoc settings.

Having shown that the proportional fairness is an appropriate optimization objective for our problem, in the second part of the thesis we consider a joint optimization of rates, transmission powers, medium access (scheduling) and routing, where the goal of the optimization is to achieve proportional fairness. We first analyze networks built on physical layers that have a rate which is

a linear function of SNR at the receiver (such as UWB or low-gain CDMA systems). We find that the optimal solution is characterized by the following principles: (1) Whenever a node transmits, it has to transmit with the maximum power; otherwise it has to remain silent ($0 - P^{MAX}$ power control). (2) Whenever data is being sent over a link, it is optimal to have an exclusion region around the destination, in which all nodes remain silent during transmission, whereas nodes outside of this region can transmit in parallel, regardless of the interference they produce at the destination. (3) When a source transmits, it adapts its transmission rate according to the level of interference at the destination due to sources transmitting in parallel. (4) The optimal size of this exclusion region depends only on the transmission power of the source of the link, and not on the length of the link nor on positions of nodes in its vicinity. As for the routing, we restrict ourselves to a subset of routes where on each successive hop we decrease the distance toward the destination. We also show that (5) relaying along a minimum energy and loss route is better than using longer hops or sending directly, which is not obvious since we optimize rate and not power consumption. Finally (6), the design of the optimal MAC protocol is independent of the choice of the routing protocol. We present a theoretical proof of optimality of $0 - P^{MAX}$ power control, and the remaining findings we show numerically on a large number of random network topologies.

Next, we consider narrow-band networks, where rate function is a strictly concave function of SNR. There, previous findings do not always hold. We show that in some cases, the size of the exclusion region and the optimal routing depend on transmission powers, and that the optimal MAC design depends on the choice of routing. Nevertheless, as we show with the example of 802.11 networks, a significant improvement over the existing 802.11 MAC can be achieved even with simpler, suboptimal strategies. Although this result is shown by simulations on a simplified model, it still gives further directions on how to improve the performance of RTS/CTS based protocols.

Version Abrégée

Nous nous intéressons à la conception multi-couche de réseaux sans fils auto-organisés. Avec une approche traditionnelle, la couche physique, celle d'accès au canal et la couche de routage sont traitées et optimisées de façon disctinctes, ce qui tend à entraîner la création d'architectures largement sous optimales. Nous proposons au contraire d'optimiser ces trois couches de manière jointe et nous démontrons que cette approche apporte un gain de performance significatif par rapport à une approche conventionelle.

Dans la première partie de ce travail, notre objectif est de sélectionner une mesure de performance appropriée pour le problème d'optimisation sous-jacent. A cette fin, nous analysons plusieurs mesures de performance existantes pour les réseaux sans fils auto-organisés telles que: maximisation de la somme des débits, équité max-min et équité proportionelle.

Premièrement, nous démontrons à travers plusieurs examples que la définition d'équité max-min ne va pas de soit pour tous les examples. En effet, il existe des exemples pour lesquels l'équité max-min ne peut être formellement définie. Nous exhibons la classe des ensemble d' allocations de débits atteignables pour laquelle une équité max-min est définie. De plus, nous montrons que cette classe comprend les ensembles de débits atteignables pour tous les cas de réseaux auto-organisés. Nous décrivons également un algorithme centralisé permettant de calculer les allocations de débits satisfaisant une équité max-min lorsqu'elles existent.

Deuxièmement, nous comparons efficacité et équité des trois mesures de performance sélectionnées dans le cas de réseaux sans-fils auto-organisés. De manière similaire aux réseaux cablés, maximiser la somme des débits entraîne une large perte d'équité ainsi qu'une famine importante pour tous les liens avec un canal de mauvaise qualité. Dans le cas de l'équité max-min, tous les liens obtiennent un débit équivalent provoquant une perte d'efficacité significative. Il est intéressant de noter que ce résultat est largement différent de celui obtenu dans le cas d'un réseau cablé. Cet ensemble de résultats offre des explications théoriques à plusieurs phénomènes observés dans le cas de réseaux sans-fils 802.11 ou à large bande. Pour terminer cette première partie, nous montrons qu'une équité proportionelle permet d'atteindre un juste milieu entre efficacité et équité

et que cette dernière est une bonne candidate comme mesure de performance pour des réseaux sans-fils auto-organisés.

Dans la seconde partie de ce travail, nous nous attachons à optimiser de façon jointe dans un réseau sans-fils auto-organisé les débits, puissances de transmission, instants d'accès au canal et routage en utilisant comme mesure de performance l'équité proportionelle précedemment sélectionnée. Nous commençons avec des réseaux construits sur des couches physique dont le débit est une fonction linéaire du rapport signal sur bruit au récepteur (par exemple avec une couche physique à large bande ou alors une couche physique "CDMA" à faible gain).

La solution optimale obtenue est caractérisée par les principes suivants: (1) Lorsqu'un noeud transmet, c'est avec la puissance maximale (contrôle de puissance $0 - P^{MAX}$) (2) Lorsqu'un noeud envoit des données, il est optimal d'avoir une région d'exclusion autour de la destination. Tous les noeuds présent à l'intérieur de cette région d'exclusion doivent rester silencieux lors de la transmission. A l'inverse, les noeuds à l'exterieur de cette région d'exclusion peuvent transmettrent en parallèle quelque soit le niveau d'interférence qu'ils créent. (3) Lorsqu'une source transmet, elle adapte son débit au niveau d'interférence mesuré à la destination. (4) La taille optimale de la région d'exclusion ne dépend ni de la distance entre la source et la destination, ni de la position des noeuds voisins de la destination mais uniquement de la puissance de transmission. En ce qui concerne le routage, nous nous restreignons à un sous-ensemble de routes ou à chaque relais, la distance à la destination est diminuée. Nous démontrons également que (5) le relayage en utilisant la route qui minimise l'énergie et les pertes est plus efficace que de favoriser des longs relais ou bien encore que de transmettre directement à la destination. Finalement (6), la conception du protocol optimal d'accès au canal est indépendent du choix du protocol de routage. Nous présentons également une preuve de l'optimalité du contrôle de puissance $0 - P^{MAX}$ et vérifions nos résultats de façon numérique sur une large palette de topologies.

Pour terminer, nous traitons le cas des réseaux a bande étroite où le débit est une fonction strictement concave du rapport signal sur bruit. Dans ce cas, les résultat presentés auparavant ne sont pas toujours valides. Nous montrons que dans certains cas, la taille de la région d'exclusion peut dépendre de la position et de la puissance de transmission des sources concurrentes. De même il se peut que l'architecture optimale d'accès au medium tienne compte du protocole de routage utilisé. Néanmoins, à travers un example utilisant un réseau sans-fil 802.11, nous montrons que des améliorations substantielles peuvent être atteintes en utilisant certaines stratégies sous-optimales. Quand bien même ces résultats sont obtenus par simulation avec des modèles simplifiés, ils indiquent les directions à prendre pour améliorer la performance des réseaux de type 802.11.

Acknowledgements

First and foremost I want to thank Jean-Yves for his excellent guidance through the labyrinths of research, helping me to clear my thoughts and grasp problems from the right sides.

I would also like to thank to the Ultra-Weird Bunch, Ruben and Jörg, for the wonderful time we had working together, and for supporting each other through sleepless nights of paper-submission deadlines.

Many thanks to all the colleagues and friends with whom I shared an office, a basement or a lab: Aleddine, Catherine, Chadi, Chirdeep, Emre, Etienne, Felix, Gianluca, Hung, Imad, Jacques, Jean-Pierre, Jochen, Jun, Ljubica, Lukas, Maciej, Mario, Mark, Mathilde, Matthias, Maxim, Michal, Milan, Naouel, Natasa, Olivier, Patrick, Paul, Petteri, Silvia, Slavisa, Sonja and Srdjan.

Special thanks go to our secretaries, Angela, Danielle and Holly, and our system administrators, Jean-Pierre, Marc-André and Phillipe, who quickly made all the unimportant problems inexistent.

I also thank the colleagues with whom I spent a wonderful and fruitful six months at IBM Zurich: Beat, Gero, Linh, Martin, Pierre, Walter and Paul. I am grateful to Emre Telatar for making time and having patience to discuss and teach me different topics in numerous meetings.

Many thanks to my neighbors, Bojana, Jelena, Maja, Mina, Deki, Igi, Misko, and all the other friends that made Lausanne a very enjoyable place to live.

Finally, I am immensely indebted to my parents and my brother for their love and support throughout my everlasting studies, and for the thirst for knowledge they infected me with.

Contents

1 Introduction **1**

2 Related Work **13**

2.1 Performance Objectives 13

2.2 Cross-Layer Design 16

2.3 Discussion 21

3 A Unified Framework for Max-Min and Min-Max Fairness with Applications **23**

3.1 Introduction 23

3.2 Motivation 25

3.2.1 Max-min Fairness 25

3.2.2 Microeconomic Approaches to Fairness 25

3.2.3 Bottleneck and Water-Filling 26

3.2.4 When Bottleneck and Water-Filling Become Less Obvious 28

3.2.5 When Bottleneck and Water-Filling Do Not Work 30

3.3 Max-Min and Min-Max Fairness in Euclidean Spaces 31

3.3.1 Definitions and Uniqueness 31

3.3.2 Max-Min Fairness and Leximin Ordering 32

3.3.3 Existence and Max-Min Achievable Sets 33

3.3.4 Max-min Fairness as A Limiting Case of Utility Fairness 36

3.4 Max-Min Programming and Water-Filling 39

3.4.1 The Max-Min Programming (MP) Algorithm 39

3.4.2 The Water-Filling (WF) Algorithm 42

3.4.3 Complexity Of The Algorithms In Case Of Linear Constraints 44

3.5 Example Scenarios 45

3.5.1 Minimum Energy Wireless Network 45

3.5.2 Load Distribution In P2P Systems . . . 47
3.5.3 Maximum Lifetime Sensor Networks . . . 48
3.6 Summary of Main Findings and Conclusions . . . 51

4 Rate Performance Objectives of Multi-hop Wireless Networks 53
4.1 Introduction . . . 53
4.2 Rate-based Performance Metrics with Power Constraints . . . 54
4.2.1 The Tension Between Efficiency and Fairness . . . 55
4.2.2 Utility Fairness . . . 56
4.3 System Assumptions and Modelling . . . 56
4.3.1 Physical Model Properties . . . 57
4.3.2 MAC Protocol . . . 58
4.3.3 Routing Protocol and Traffic Flows . . . 59
4.3.4 Power Constraints . . . 60
4.4 Mathematical Formulation of The Feasible Sets and of the Metrics . . . 61
4.4.1 Feasible Set of Rates . . . 61
4.4.2 Feasible Set of Transport Rates . . . 63
4.4.3 Design Criteria . . . 63
4.4.4 Performance Indices . . . 64
4.4.5 Performance Metrics . . . 65
4.5 Max-min Fairness . . . 66
4.5.1 Solidarity Property and Equality . . . 66
4.5.2 Solidarity of The Feasible Rate Set of A Multi-hop Wireless Network . . . 67
4.5.3 Equality of Max-min Fair Rates . . . 69
4.5.4 Influence of Long-Term Average Power Constraint . . . 71
4.5.5 An Application to an 802.11 Network . . . 71
4.5.6 When Max-min Fairness Does Not Lead To Equality . . . 72
4.6 Maximizing Total Capacity . . . 72
4.6.1 Asymptotic Results . . . 72
4.6.2 Numerical Results . . . 76
4.7 Utility Fairness . . . 80
4.8 Influence of Long-Term Average Power Constraint . . . 83
4.9 Summary of Main Findings and Conclusions . . . 84

5 Optimal Power Control, Rate Control, Scheduling and Routing in UWB Networks 87
5.1 Introduction . . . 87
5.2 Motivation and Formulation of The Problem . . . 89
5.2.1 Physical Layer Properties . . . 89
5.2.2 Cross-Layer Design in Wireless Networks . . . 90
5.2.3 Performance Metric . . . 90
5.2.4 Routing and Mobility . . . 91
5.2.5 Problem Formulation . . . 92
5.3 Assumptions and Modeling . . . 92
5.3.1 Notations . . . 92
5.3.2 Physical Network Model . . . 92
5.3.3 MAC Layer . . . 96
5.3.4 Routing . . . 97
5.3.5 Mobility . . . 98
5.3.6 Traffic Demand and Flow Control . . . 99
5.3.7 Performance Objectives . . . 99
5.4 Mathematical Analysis of the Model . . . 99
5.4.1 Mathematical Formulation . . . 99
5.4.2 Optimal Power Allocation . . . 100
5.4.3 Minimum Energy and Loss Routes . . . 101
5.4.4 Techniques for Solving the Problem . . . 102
5.5 The Static Ring Case . . . 103
5.5.1 Proportional Fairness . . . 104
5.5.2 Scheduling . . . 104
5.5.3 Power Allocation . . . 105
5.5.4 Routing . . . 106
5.5.5 Exclusion Regions . . . 107
5.6 Mobile Networks in A Plane . . . 108
5.6.1 Routing Strategies . . . 110
5.6.2 Scheduling Strategies . . . 110
5.7 Numerical Results . . . 112
5.7.1 An Example Scenario . . . 112
5.7.2 Numerical Values of System Parameters . . . 114
5.7.3 Homogeneous Networks with Homogeneous Traffic . . . 115

5.7.4 Non-Homogeneous Networks . . . 116
5.7.5 The Effects of Mobility and The Cost of Routing . . . 118
5.8 Application to Very Low-Power UWB Networks . . . 120
5.8.1 Interference Mitigation . . . 122
5.9 Summary of Main Findings and Conclusions . . . 125

6 Power Control is Not Required for Wireless Networks in the Linear Regime 129
6.1 Introduction . . . 129
6.2 Motivation . . . 130
6.2.1 Power Control and Optimal Wireless MAC Design . . . 130
6.2.2 Rate Adaptation and Rate Function . . . 130
6.2.3 Linear Regime . . . 131
6.2.4 Rate Maximization and Power Minimization . . . 131
6.2.5 Power Control in Existing Systems . . . 131
6.2.6 Performance Comparison . . . 133
6.3 System Assumptions and Modeling . . . 133
6.3.1 MAC Protocol . . . 134
6.3.2 Routing Protocol and Traffic Flows . . . 134
6.3.3 Power and Rate Constraints . . . 134
6.3.4 Performance Objectives . . . 135
6.4 Mathematical Model . . . 136
6.4.1 Notations . . . 136
6.4.2 Mathematical Formulation . . . 136
6.5 Main Findings . . . 138
6.5.1 Rate Maximization . . . 138
6.5.2 Power Minimization . . . 141
6.5.3 Numerical Example . . . 141
6.6 Summary of Main Findings and Conclusions . . . 143

7 Cross Layer Design for Narrow-Band Networks 145
7.1 Introduction . . . 145
7.2 Assumptions and Modeling . . . 147
7.2.1 Physical Layer . . . 147
7.2.2 Optimization Problem . . . 148
7.2.3 Network Topologies . . . 148

7.3 The Ring Case . . . 149
7.3.1 Analysis . . . 149
7.3.2 Results . . . 150
7.4 Networks in Plane . . . 158
7.4.1 General Findings . . . 158
7.4.2 Application to 802.11 . . . 159
7.4.3 Other Examples . . . 161
7.5 Summary of Main Findings and Conclusions . . . 161

8 Conclusions **165**

Bibliography **169**

A Receiver Architectures for Transmit-Only, UWB-Based Sensor Networks **177**
A.1 Introduction . . . 177
A.2 System Model and Performance Objectives . . . 179
A.2.1 Sensor Node Architecture . . . 179
A.2.2 Ultra Wideband Physical Layer . . . 179
A.2.3 Cluster Head Design Objectives . . . 179
A.3 Detailed Description of The Proposed Architectures . . . 181
A.3.1 Simple CHR Architecture . . . 181
A.3.2 Adaptive-Threshold CH Architecture . . . 181
A.3.3 Switched CH Architecture . . . 186
A.4 Performance Comparison . . . 187
A.5 Conclusions . . . 188

B Curriculum Vitae and List of Publications **191**

List of Figures

3.1 An example of a network with its feasible rate set, and max-min allocation obtained by water-filling . . . 27
3.2 More examples of feasible rate sets . . . 28
3.3 A simple multi-path example . . . 29
3.4 When water-filling does not work . . . 30
3.5 Examples of 2-dimensional sets that do not have max-min fair allocation . . . 33
3.6 Choice of constants for the proof of Theorem 3.1 . . . 35
3.7 Choice of constants for the proof of Theorem 3.2 . . . 37
3.8 Minimum energy networks . . . 46
3.9 Sensor example . . . 50

4.1 Two examples of schedules . . . 58
4.2 Illustration of the solidarity property . . . 67
4.3 A random network and corresponding rate distributions . . . 70
4.4 The fairness and efficiency indices of optimal rates with respect to different performance objectives . . . 78
4.5 Non-uniform topologies . . . 80
4.6 Random fading and noise power . . . 81
4.7 Efficiency index and fairness index of the rates that maximizes total capacity and transport capacity , and the proportionally fair ($\xi = 1$) rates . . . 82
4.8 Efficiency index and fairness index of max-min and proportional fairness for finite long-term average power constraint . . . 83

5.1 Ultra-wide band physical layer with PPM, the model of Win Scholtz . . . 93
5.2 Different routing policies . . . 98
5.3 Ring topology . . . 103
5.4 Performances of different routing and scheduling strategies for ring topologies . . . 106

5.5 Illustration of Proposition 5.6 . . . 107
5.6 Short representations of the optimal policies for rings . . . 108
5.7 An example of the exclusion region . . . 111
5.8 An example of a network with 50 randomly distributed nodes and 25 randomly distributed flows. . . . 113
5.9 Histogram of flow rates achieved on the network from Figure 5.8 with MELR routing and different schedulings . . . 113
5.10 Histogram of flow rates achieved on the network from Figure 5.8 using scheduling strategy 2 and MELR or DIR routing . . . 114
5.11 Utilities of different routing and scheduling algorithms applied on homogeneous random networks and different maximal power constraints . . . 116
5.12 Utilities of different routing and scheduling algorithms applied on random networks with non-uniform node distribution . . . 117
5.13 Utilities of different routing and scheduling algorithms applied on homogeneous random networks with base-stations . . . 117
5.14 Utilities of different routing and scheduling algorithms applied on homogeneous random networks with non-uniform maximal power constraints . . . 118
5.15 Utilities of different routing algorithms applied on mobile homogeneous random networks and different maximal power constraints . . . 119
5.16 Cylindric multiple interferer scenario . . . 123
5.17 Rates achieved in the cylindric scenario . . . 124

6.1 A simple example of a network with 2 links . . . 142

7.1 Analyzed topologies: ring and line . . . 148
7.2 Illustration of Proposition 7.3 . . . 151
7.3 Short representations of the optimal policies for $n = 8$ and $n = 18$ . . . 151
7.4 Maximum rates for DIR and MER routing . . . 152
7.5 Transition point from MER to DIR as a function of power . . . 153
7.6 Optimal scheduling for different power constraints . . . 153
7.7 Number of nodes active in one time slot . . . 154
7.8 Rates per distance and transition power constraints . . . 155
7.9 Optimal scheduling in the line case . . . 156
7.10 Number of nodes transmitting at the same time . . . 157

7.11 Fraction of nodes that are active in a same slot and the average distance between them 158
7.12 Performance comparison of different MAC protocols on random networks 163

A.1 An illustration of a transmit-only sensor network in an intrusion detection scenario 180
A.2 Illustration of the detection threshold principle 182
A.3 Simulation results for a network with 50 SNs 184
A.4 An example of packet arrivals that illustrates the adaptation mechanism 185
A.5 Aggregate utilization as function of the total load 187

Chapter 1

Introduction

Dissertation Overview

Motivation

Traditionally, network protocols are divided into several independent layers. Each layer is designed separately, and the interaction between layers is performed through a well-defined interface. The main advantage of this type of approach is architectural flexibility. One implementation of a layer can be seamlessly replaced with another implementation. For example, if an old protocol has to be replaced with a newer one, there is no need to modify the rest of the network stack.

Apart from the flexibility, in wired networking applications this approach to physical, MAC and network layers does not impose a performance penalty: separate wired links are independent of each other, and do not affect each other's performance.

The story is different for wireless networks. One of the differences is that links can no longer be viewed as separate entities whose performances are independent of each other. Each wireless transmission can be heard by nodes in the neighborhood and is perceived by them as interference. As a consequence of the interdependency, there is a need for a more complex medium access mechanism. On the one hand, this mechanism should be able to control the amount of interference experienced by receivers. On the other hand, it should exploit spatial reuse and, in certain cases, enforce concurrent transmissions, in order to maximize the performance.

Since medium access protocol controls the amount of interference in the network, it influences the performance of the physical layer. If the total amount of the interference at a receiver during a reception of a packet is high, the physical layer should decrease the transmission rate to

cope with it. On the contrary, if the interference is low, the physical layer should benefit from the conditions and transmit with a high rate. This is again contrary to the approach in a wired network where any two concurrent transmissions always cause a collision, and there is no concept of rate adaptation as a function of existing traffic.

Another degree of freedom that does not exist in the physical layers of wired networks is power control. The higher the packet transmission power is, the higher the received power is, hence the higher the transmission rate. However, if we increase the transmission power, at the same time we increase the interference to other nodes. Power control is thus tightly coupled with both the physical and medium access layers.

Perhaps the least obvious interdependency is between the physical layer, medium access and routing. Wireless medium allows connections between any two nodes in a region: the further away the nodes are, the lower the achievable communication rate will be. If the direct connection can be establish only with a very low rate, the routing protocol may decide to relay over intermediate nodes. This way, a single packet will be transmitted on several short links: it will consume the resources of several nodes, but for a shorter period of time, since each intermediate transmission will be on a higher rate.

It is easy to see that a change in routing policy may change the performance of the layers underneath. Namely, if a routing protocol decides to relay one flow over several hops, instead of transmitting directly from a source to a destination, we will have a number of short links on the route. Some of these links will be active at the same time, and this will create more interference on other nodes than a single long hop. At the same time, these shorter links will be more resistant to interference than the single long link.

From the above examples it is clear that a change in a protocol of one layer will affect the performance of other layers. The independent design of different layers may yield a grossly suboptimal network. The goal of this thesis is to analyze a cross-layer network design that will yield jointly optimal physical, medium access and routing layers for wireless ad-hoc networks.

Performance Objectives

A cross-layer design of a wireless ad-hoc network is essentially an optimization problem with several degrees of freedom. Before proceeding to solving the optimization problem, we have to define the performance objective.

Roughly, performance objectives, in the case of wireless ad-hoc networks, can be divided in two categories. The first category comprises power-based performance objectives that maximize network lifetime. A typical example is a sensor network. Traffic requirement is low, and the

main goal is to maintain a network in operation as long as possible.

In the other category, there are rate-based performance objectives. For these, the goal is to maximize flow rates. Typical examples are wireless LANs, networks of computer peripherals (e.g. wireless replacements for USB or FireWire) or home appliances. The overall power consumption of nodes of such a network is typically much larger than the energy spent on wireless transmissions, hence there is no need to optimize network lifetime, whereas applications, such as video, audio or data transmissions, may require high data rates.

In this thesis we consider wireless networks with best-effort traffic and we focus on rate-based performance metrics. We analyze the well-known rate-based performance metrics, originally defined in the wired networking context: rate maximization, proportional fairness and max-min fairness, and some of there modifications. We first focus on max-min fairness.

A Unified Framework for Max-Min Fairness

In the context of wired networks, max-min fair rate allocation is defined to be the flow rate allocation in which every flow has a bottleneck link [8]. This rate allocation can be obtained using the well-known water-filling algorithm. However, it is not always obvious how to generalize the notion of a bottleneck link and the water-filling approach to an arbitrary problem. A simple example is a wireless ad-hoc network. Even if every flow has a bottleneck link, the allocation may not be max-min fair. In particular, by decreasing the transmit power of some links, one can decrease the interference on the receiver of one of the bottleneck links and thus increase the link's rate. This way, a flow will loose its bottleneck, and one might be able to increase its rate. It is difficult to define the concept of a bottleneck link and of the water-filling in the given example, and furthermore, it is not obvious if the max-min fair rate allocation can be defined at all in this example.

We propose a unified framework for max-min fairness. We extend the existing concept of max-min fair allocation on sets in Euclidean spaces and we show that it exists for a large class of sets. This class of sets includes (but is not limited to) all convex sets. In the previous example it can be shown that the set of all feasible rate allocations that can be achieved (using different power adaptation, rate adaptation, medium access and routing strategies), is a convex set. It then follows from our framework that there exists a single max-min fair allocation on that set. We give several different examples where it is not obvious that a max-min or min-max fair allocation exists, and we show the existence by analyzing properties of the feasible rate sets.

Next we present a general-purpose, centralized algorithm called max-min programming, and we prove that it finds the max-min fair allocation in all cases where it exists. Its complexity

is of the order of N linear programming steps in R^N, in general, whenever the feasible set is defined by linear constraints. Furthermore, we recall the definition of the free disposal property [54] and show that, whenever it holds, max-min programming degenerates to the simpler water-filling algorithm [8], whose complexity is much less. The free-disposal property corresponds to the cases where a bottleneck argument can be made, and water-filling is the general form of all previously known centralized algorithms for such cases.

Efficiency and Fairness of Rate-Based Performance Metrics

Once we have shown that the max-min fair rate allocation can be defined in our setting, we analyze the properties of the three design objectives: maximizing the sum of rates, proportional fairness and max-min fairness, within the framework of wireless ad-hoc networks.

One of main reasons that maximizing the sum of rates is a widely used objective in ad-hoc network design is its simplicity. However, it has already been observed in the context of wired networking that maximizing the sum of rates exhibits a large unfairness [8], and the same phenomenon has been confirmed in wireless settings [21]. One way to remedy the unfairness is to consider maximizing a weighted sum of rates, where the importance of each flow can be tuned through the corresponding weight. A famous example of such an approach is the transport capacity metric [30]. It is defined as the sum of the rates of flows multiplied by the distance over which a flow carries information. Another way to remedy unfairness is to consider different objectives that are constructed in a way that guarantees fairness, like max-min fairness or utility fairness.

We show that both maximizing the sum of rates and the transport rates metrics are unfair metrics in wireless setting. We prove that, for both metrics, when transmit power goes to infinity, all but the shortest flows will have zero rates. We also demonstrate that the phenomenon occurs in a large number of realistic examples.

Next, we consider max-min fairness. Although it is perfectly viable for a wired network, it is much less so in our setting. We show that, in the limit of long battery-lifetimes, the max-min allocation of rates always leads to strictly equal rates, regardless of the MAC layer, network topology, channel variations, or choice of routes and power constraints. This is due to the "solidarity" property of the set of feasible rates. This results in all flows receiving the rate of the worst flow, which leads to severe inefficiency. We show numerically that the problem persists when battery-lifetime constraints are finite. This generalizes the observation reported in the literature [7] that, in heterogeneous settings, 802.11 allocates the worst rate to all stations and shows that this is inherent to any protocol that implements max-min fairness.

Utility fairness is an alternative to max-min fairness. It approximates rate allocation performed by TCP in the Internet. We analyze by numerical simulations different utility functions and we show that the proportional fairness [47] of rates or transport rates (a particular instance of utility based metrics) is robust and achieves a good trade-off between efficiency and fairness, unlike total rate or maximum fairness. We thus recommend that metrics for the rate performance of mobile ad-hoc networking protocols be based on proportional fairness.

Optimal Cross-Layer Design

In the second part of this thesis we analyze the optimal cross-layer design of a wireless ad-hoc network. The main building blocks of the wireless network design are power adaptation, medium access (or scheduling), rate adaptation and routing. The role of power adaptation is to decide upon the transmission power of every transmitted packet. Medium access organizes concurrent transmissions in order to control the level of interference in the network. Rate adaptation is performed on each link separately and its goal is to adapt the rate of a transmission to the level of interference at the receiver. Finally, the role of routing is to choose the optimal relaying strategy. Our goal is thus to jointly optimize all these building blocks.

In the first part of the thesis we show that the proportional fairness is an appropriate rate-based performance metric for wireless ad-hoc network. We will use it as the optimization objective for the joint optimization problem.

We formulate the joint optimization problem on a general model of a network that incorporates all degrees of freedom defined by protocols. This is a convex optimization problem that always has a solution. However, it is very complex and has previously been solved only for networks with a small number of (less than 10) nodes.

One of the parameters of our model is a rate function that defines the maximum achievable rate on a link as a function of the signal-to-interference-and-noise (SINR) level at the receiver. This function is typically a concave function of SINR. In the case of physical layers with a large bandwidth and/or small transmission powers, the rate function becomes a linear function of SINR. We first consider networks with linear rate functions (such as UWB or low-gain CDMA systems).

When the rate function is linear, we show that the only optimal power control is $0 - P^{MAX}$: whenever a node is transmitting, it has to transmit with the maximum power; otherwise it should remain silent. Any power control different from $0 - P^{MAX}$ is strictly suboptimal. We prove analytically that this result holds for an arbitrary network with an arbitrary routing protocol and with the optimal medium access and rate adaptation. Furthermore, we show that this result does

not hold only for the proportionally fair rate allocation but for any Pareto efficient rate allocation. We also extend this result to energy conservation scenarios. There, we consider networks with minimum rate guarantees, and we search to minimize average power consumption. We show that even in this case, $0 - P^{MAX}$ is one of the optimal power control strategies (but not necessarily the only one).

A medium access has to manage the level of interference experienced by each active destination. Whenever data is being sent over a link, there should be an exclusion region around the destination, in which all nodes remain silent during transmission, whereas nodes outside of this region can transmit in parallel, regardless of the interference they produce at the destination. In the optimal case, this approximately corresponds to having all exclusion regions of the same radius; the radius depends only on transmission powers, not on the link sizes or positions of other nodes. We show by heuristics that this strategy is very close to optimal.

The optimal rate adaptation strategy is that every source, when transmitting, should adapt its transmission rate according to the level of interference at the destination due to sources transmitting in parallel.

As for the routing, we restrict ourselves to a subset of routes where on each successive hop we decrease the distance toward the destination, and we show that relaying along a minimum energy route is better than using longer hops or sending directly, which is not obvious since we optimize rate and not power consumption. We also show that the optimal power control, medium access, and rate adaptation do not depend on the choice of the routing protocol. A consequence of this is that in the optimal cross-layer design medium access and physical layer should be jointly designed, whereas routing can be incorporated as an independent layer, like it is done in wired architectures. We show this finding by heuristics and numerical simulations on a large number of random network topologies.

We demonstrate our findings by applying them on a low-power pulse-based ultra-wide band (UWB) network, based on the Win-Scholtz physical layer model [96]. This physical layer has a linear rate function, hence our findings directly apply. We first show that the size of the optimal exclusion region for this network is of order of 2m.

We then introduce an interference mitigation technique on the physical layer, in order to further decrease the exclusion region. In a pulse-based UWB network, it is optimal to use short, strong pulses. Since we focus on low-power networks, the delay between two consecutive pulses is very long, hence the probability that a data pulse from a source and a pulse from an interferer transmitting in parallel will collide at a destination is small. Nevertheless, if the two pulses collide and the interfering pulse is much stronger, then its impact on the decoding process will be

huge since it will influence a decision on a potentially large codeword. Our interference mitigation technique consists in monitoring and detecting a collision with a very strong interfering pulse. Then, instead of passing the corresponding sample to the decoder, we declare an erasure.

By adding the interference mitigation to the physical layer, we modify its rate function. We show by simulations that the size of the optimal exclusion region is now very small (of the order of tens of centimeters). Therefore, an optimal medium access protocol should not perform any exclusions and should let all nodes transmit whenever they have data to transmit, in parallel to all other existing transmissions.

The theoretical analysis presented above has an impact on the MAC and PHY design paradigm for low-power UWB networks. The first implication is that there is no need for a distributed mutual exclusion protocol like CSMA/CA or Aloha. It is optimal for all nodes to transmit whenever they have data to transmit. Also, there is no need for power control, and all transmissions are performed with the maximal power. Instead, the complexity of the MAC is in the rate adaptation. Each source has to adapt its communication rate to the SNR at a destination. However, unlike mutual exclusion, this is a purely local task and significantly easier to implement. For example, there is no hidden terminal problem (which is solely due to imperfect information exchange in distributed MAC protocols). Another remaining protocol issue is how to resolve the contention of several transmitters for the same destination. But this is also a local problem, which is considerably simpler to solve.

A protocol designed along these guidelines is DCC MAC [58]. It is shown in [58] that DCC MAC outperforms, by far, contemporary MAC protocols. In addition, unlike the other protocols, it does not degrade performance when a network is mobile, or when TCP is used over it, due to its inherently decentralized nature.

We next consider networks with physical layers with rate functions that are concave but non-linear. We first inspect a case where the rate is log of SINR, which corresponds to a single-user Gaussian channel. We consider a simple symmetric network on a ring where all nodes are equally spaced, and each node talks to a neighbor n-hops away on the right. The symmetry of the network facilitates calculations, and we can derive very precise heuristics that describe the optimal solution. We find that in this example, some of the properties of the optimal solution in networks with linear rate functions do not hold. Specifically, the size of the exclusion region does not depend only on transmission powers but also on link sizes and positions of other nodes. The optimal routing depends on transmission powers and does not always correspond to the minimum energy route. Finally, the size of the exclusion region depends on the choice of routing, and a different medium access control is needed for different routing protocols.

From the simplified 1D example we learned that the optimal design is not the same for different types of rate functions. Furthermore, in the case of non-linear rate functions, it is difficult to derive any exact characterization of the optimal design for an arbitrary network. Instead, we try to derive some heuristics, and we focus on networks with 802.11 PHY, which is one of the most widely used narrow-band physical layers. We set all nodes to use maximum power when transmitting, we use exclusion regions of fixed sizes, and we centrally calculate the optimal routes. Although the obtained heuristics are not necessarily optimal, we show that they outperform the existing 802.11 MAC protocol. The main reason is that the size of the exclusion region in 802.11 MAC is too large. Exclusion regions in 802.11 MAC are implemented via RTS/CTS. Before a node sends or receives a packet, it will send an RTS or CTS control packet, respectively. All nodes in the neighborhood that hear RTS or CTS will cease transmitting. However, these control packets are sent with the maximum power. Hence, too many nodes in the neighborhood are excluded and spatial reuse is low. We show that, by simply reducing the range of RTS/CTS in 802.11 MAC and by dynamically adapting the rate of a link to the interference caused by concurrently transmitting nodes outside of the exclusion region, we can increase link rates by 30%-60%.

Finally, in the appendix, we discuss a wireless network design within a different framework. We are interested in designing a low-cost, ultra-wide-band (UWB) sensor network that consists of a large number (in the order of 100) sensor nodes (SNs). These SNs sense the environment and transmit the resulting data via a wireless radio link to so-called cluster heads (CHs); the number of CHs in the sensor network is in the range of 1 to 10. A physical layer with favorable robustness and relatively simple transmitters is pulse based UWB-radio. However, the implementation of a corresponding receiver is much more complex, therefore we focus on SNs that are only equipped with a transmitter. A consequence of the transmit-only feature is that the access of the SNs to the radio channel is uncoordinated such that some packets collide, and this decreases the performance. For complexity reasons we consider simple noncoherent receivers at the CHs.

To improve the robustness to packet collision and thus, to increase the number of successfully received data packets from the SNs, we propose two novel low-complexity noncoherent receiver architectures: (i) The adaptive CH receiver (CHR), which maintains an adaptive threshold; only packets whose signal strength is larger than the threshold are received. (ii) The switching CHR, with an additional acquisition circuit. This receiver allows for switching the focus from a packet of a given SN to a packet from another SN with higher signal strength. We show by simulation that the number of received packets during burst periods can be drastically increased with the adaptive CHR, and even more with the switching CHR.

Goals and Achievements

The main goal of this thesis is to find an optimal cross-layer design of a multi-hop wireless network. We solve the problem by dividing it into three sub-problems, that we solve subsequently:

- Identify main building blocks of a wireless network and model each of them as a set of free variables. Using these models we formulate a cross-layer design as an optimization problem.
- Find an appropriate rate-based design objective for multi-hop wireless networks. Analyze the existing rate maximization performance metrics and show which ones are applicable to ad-hoc wireless network settings.
- Find the jointly optimal physical, medium access and routing layer that maximizes an appropriate rate-maximization performance metrics, in order to the understand the principles of cross-layer design of wireless ad-hoc networks.

The primary achievements of this thesis are:

- A general framework that shows the existence of max-min and min-max fairness on a large class of sets in Euclidian spaces, and an algorithm that finds the optimal allocations whenever they exist.
- Fairness and efficiency analysis of the existing rate maximization metrics applied on wireless ad-hoc networks; we show that maximizing the total sum of rates is unfair, max-min fairness is inefficient, and proportional fairness makes a good trade-off between efficiency and fairness.
- Characterization of the optimal cross-layer protocol, which achieves proportional fairness, for wireless networks with linear rate function:
  - In the optimal protocol, whenever a node is transmitting, it should transmit with a full power; otherwise, it should remain silent ($0 - P^{MAX}$ power control).
  - There should be an exclusion region around each destination whenever that destination is receiving.
  - An approximately optimal exclusion region has a radius that is independent of positions of adjacent nodes and link size, and it depends only on the source's transmission power.

  - Relaying along a minimum energy and loss route is better than using longer hops or sending directly
  - The optimal routing protocol is independent of the choice of MAC protocol and vice verse

- We demonstrate an application of our findings on a low-power ultra-wide band network. We show that with an interference mitigation technique added to the physical layer, the optimal size of the exclusion region becomes very small, hence there is no need for exclusions.

- When the physical layer is narrow-band, we show that the last three findings from the wide-band case do not hold. We then consider the rate function of 802.11 physical layer, as a particular case of a narrow-band physical layer. 802.11 MAC essentially enforces exclusion regions through RTS/CTS protocol. We find that RTS/CTS are sent with too high powers, the corresponding exclusion regions are too large, hence there is not enough spatial reuse. We show that by decreasing exclusion region size we can increase rates from 30% - 60%.

- In the appendix, we discuss the design of a transmit-only sensor network. We propose two new architectures of cluster heads (devices that collect packets from sensors) that improve a number of successfully received packets.

What This Thesis Is Not About

- We characterize the optimal solution, in parts exactly, and in parts using heuristics. We do not find the optimal solution for any given network topology (this is still a complex problem we do not know how to solve).

- We do not design an actual protocol. We put forward a mathematical model of a network and solve an optimization problem on the model. The solution to the problem gives us guidelines on how to design optimal network protocols. However, the actual protocol design is not covered in this thesis and remains for future work.

- In the optimization problem we do not consider multi-path routing. We focus only on single-path routing, the one implemented in the current wireless routing protocols. We assume all nodes use the same routing strategy.

Outline

In Chapter 2 we present an overview of the existing works that are related to our field of study. We position our research with respect to the state of the art in this field.

In Chapter 3 and Chapter 4 we discuss different design objectives. In Chapter 3 we analyze if max-min fairness can be applied in the scenarios of our interest. In Chapter 4 we analyze different properties of existing rate-based performance metrics, and identify suitable objectives for network design.

In Chapter 5 we consider cross-layer design of ultra wide-band systems. We formulate and solve the optimization problem and derive guidelines for the optimal protocol design.

In Chapter 6 we discuss the power control issues in wide-band network design. We extend findings about power control from Chapter 5 to arbitrary rate-based or energy-based metrics.

In Chapter 7, we analyze the optimal cross-layer design of narrow-band networks. We show that it is different from wide-band design. We also show how the existing 802.11 MAC can be improved using our design approach.

In Chapter 8 we present our conclusions.

In the appendix, we discuss receiver architectures for transmit-only, UWB-based sensor networks.

Chapter 2

Related Work

Our goal in this thesis is to analyze the cross-layer design of wireless networks. In the first part of the thesis, we discuss what performance objectives to use for a wireless network design. In the second part, we discuss the design guidelines that maximize the design objective. This chapter reviews the related work done in both of the fields. First we review the existing performance objectives and their implications on wireless networks. We next discuss different existing protocol designs from a cross-layer point of view. Finally, we will briefly position our research with respect to the existing works.

2.1 Performance Objectives

Performance objectives for wireless network design can roughly be divided into two categories. One category comprises energy-based performance objectives: the goal is to minimize energy dissipation of a network and increase network lifetime. Examples of this type of metric are discussed for example in [14, 77] and [12]. This category is out of the scope of this thesis, hence we will not discuss it further.

The other category comprises rate-based performance objectives: the goal is to maximize capacity. There are three common rate-based performance metrics used in networking: maximizing total capacity, max-min fairness and utility fairness.

Maximizing Total Capacity

Maximizing total capacity is a performance objective traditionally used in the design of cellular systems. This objective comes from a voice setting where all users have the same capacity need,

and the goal is to maximize the total capacity (i.e. the maximum number of calls that can be supported by a network).

It is however much less suitable for data traffic. It has been known in the world of wired networking that maximizing total capacity is an unfair design objective [8]. The same problem persists in wireless networks. In the case of cellular networks, this issue has been discussed, for example by Tse and Hanly in [90], in context of multiple-access channel. A strategy that maximizes the total capacity is such that a node with the best channel conditions in a given slot should send data. Nodes that are farther away will less frequently satisfy this constraint, but will still have a positive throughput due to the random part of fading. However, if a node is very far away from the base station, its average rate will be very small and essentially it will not be able to communicate. In [90], a remedy is found by assigning weights to node rates such that a level of fairness is assured. The implicit assumption in this type of network is that an area with mobile nodes is well covered with base stations, so there is not a great variation in distances from the mobile nodes to the closest base stations. A similar direction was taken for HDR/CDMA networks, as described in [67].

In the case of multi-hop wireless networks, variations in the distances between sources and destinations are typically much higher since a node does not talk to the closest base-station but to an arbitrary destination in a network. This makes it difficult to remedy fairness with weights, and longer flows risk low or zero throughput. For example, it has been observed in the context of Ultra Wide Band by Cuomo et. al. [21] that the unfairness of maximizing the sum of rates persists in wireless networks and some long distance flows obtain zero throughput. A famous rate-based metric based on weighted sum of rates and used in multi-hop wireless setting, is transport capacity [30].

Max-Min Fairness

Max-min fair rate allocation is defined to be the rate allocation in which no rate can be increased except at the expense of an already smaller rate (the precise definition is given in Chapter 3). Max-min fairness is originally proposed in [23]. A detailed discussion can be found in [8]. Its main purpose is to ensure fairness among different users in a network.

Max-min fairness is widely used in the framework of wireless networking. Different applications can be found in [68, 31, 78, 13, 36] and [40]. Schedulers that implements max-min fairness in wired networks have been presented in [15] and [23].

Max-min fairness is also used in wireless network design. Examples can be found in [86, 37].

Utility Fairness

A different approach to fairness in wired networking is utility fairness. Every user is assigned a utility which is a function of its rate. The goal of a network is to maximize the sum of utilities of all users. Utility fairness is explained in details in Chapter 4, and it is discussed for example in [59, 47] and [55]. By selecting utility functions, a designer can achieve different trade-offs between efficiency and fairness. This has been demonstrated in [84]. It can also be shown that TCP implements a form of utility fairness.

Variants of utility fairness are used in existing wireless multi-hop network protocols. In [63] per-link proportional fairness is considered. An algorithm that achieves the end-to-end proportional fairness using hop-by-hop congestion control is presented in [98]. In [52], a general framework for finding a utility-fair rate allocation in a multi-hop wireless network is presented.

Alternatively, this approach can be extended to combine rate-based and power-based metrics. An example, applied on UWB wireless networks, is given in [4].

802.11 Performance Objectives

In case of 802.11 networks with base stations, nodes contend for access to a base-station. The goal of the base-station is to receive an equal number of packets from each of the transmitting nodes. However, this performance objective causes a performance anomaly that was reported by Berger-Sabatel et. al. in [7] for 802.11. There, several nodes talk to a base station. One of them is far away and codes for 1 Mb/s while others are near and code for 11 Mb/s. Still, on average, all nodes achieve the same throughput of approximately 1 Mb/s. We show in Chapter 4 that this anomaly is in fact not abnormal behavior, rather a fundamental property of max-min fairness for wireless networks, regardless of any underlying physical, MAC or routing protocol.

Another type of fairness for 802.11 networks is defined in [27]. Instead of considering a fair rate sharing, Gambiroza et. al. consider a fair time sharing. They assume a physical model similar to 802.11. Two links can either transmit concurrently, or collide with each other forming a contention region. If several links compete for the same contention region, then time sharing is necessary and a form of max-min fair time sharing is proposed. Although this approach alleviates the inefficiency described in [7], it is difficult to generalize it to more general wireless physical models that include rate or power adaptations.

2.2 Cross-Layer Design

It has recently become evident that a traditional layering network approach (separating routing, scheduling, rate and power control) is not efficient for ad-hoc wireless networks [29]. This is primarily due to the interaction of links through interference, which implies that a change in power allocation or schedules on one link can induce changes in the capacities of all links in the surrounding area and changes in the performance of flows that do not pass over the modified link.

The main building blocks of a wireless network design are rate control, power control, medium access (scheduling) and routing. These building blocks are divided in layers. Typically, routing is considered in a routing layer and medium access in a MAC layer, whereas power control and rate control are sometimes considered in a PHY and sometimes in a MAC layer. The goal of this section is to review existing works and explain how each of the building blocks are implemented in different protocols, and how they are integrated in terms of cross-layer design.

Almost all protocols for multi-hop wireless networks implement a form of routing and scheduling. When it comes to rate and power control, the issue is less clear. Therefore, we classify the existing works according to the way they implement rate and power control, and for each protocol we explain how the building blocks are integrated.

We first classify the existing protocols and explain briefly each of them. Then we discuss open issues and position the work in this thesis.

Fixed-Rate, Fixed-Power Protocols

The first wireless MAC protocols for multi-hop networks were designed to control only mutual exclusion. A typical example is the original 802.11 network (whereas the latter incarnations 802.11a/b/g offered variable link rates, the original 802.11 offered communications only at 1Mbps). It always uses maximum power for transmitting a packet and aims to establish communication on a predefined link rate.

Medium access contention in 802.11 MAC is resolved by CSMA/CA and RTS/CTS packets. The combination of RTS/CTS packets and carrier sensing can be seen as a mean to enforcing **exclusion regions** around sources and destinations: nodes that hear and decode RTS/CTS packets (thus in the exclusion regions) will be excluded from transmitting in parallel, whereas nodes that do not hear RTS/CTS (thus outside of the exclusion regions) may proceed with transmitting. RTS thus generates a form of exclusion region around a source, and CTS around a destination, and all the nodes within these exclusion regions have to remain silent during the transmission. Note

that as will be discussed in Chapter 5, we need to have exclusion regions only around receivers. In 802.11 each transmitter is also a receiver since it needs to receive an acknowledgment after every transmission, hence the need to have an exclusion region around a transmitter as well.

One of the problems with the above approach is that the scheduling is done in an ad-hoc manner. It is not clear how often a link should be scheduled, and how much the performance of the network is affected by an activation of one link and the imposed silence of transmitters in the link's exclusion region. An analysis in this direction has been presented, for example, in [63]. Nandagopal et. al. define a notion of cliques. A clique represents a set of links that interfere with each other such that only one link from a clique can be active at a time. Given a network topology that define cliques, they then solve the optimization problem to find the per-link proportionally fair rate allocation.

None of the above models considers joint optimization. Rates and powers are fixed, and routing is performed independently of scheduling. One of the first models for cross-layer optimization is presented in [85]: it considers joint scheduling and routing problem, and shows that the problem can be decomposed in a pure routing and a pure scheduling problem, which are then easier to solve. An extension to this model is given in [98] where Yi and Shakkottai reuse the clique model from [63], and analyze a joint routing and scheduling problem that aims at achieving end-to-end proportional fairness.

However, the main drawback of all these approaches is that they have a very simplified interference model and they do not consider the effect of cumulative interference. Also, as we already mentioned, they do not consider a possibility of adapting rate as a function of interference, or power adaptation as a way to manage interference. Several improvements to these approaches have been proposed, and they are discussed next.

Fixed-Rate, Variable Power Protocols

There is another class of wireless network models that incorporates effects of power control. The first analysis of this type, within the framework of cellular networks, can be found in [25, 97] and [5]. Each node in a cellular network requires a minimum QoS guarantee (typically a voice conversation that requires a fixed bit rate). The QoS guarantees can then be translated into minimum signal to noise ratios. We then have the following optimization problem: given a set of links, find a power allocation that minimizes the total dissipated power, such that all links have the SNR requirements satisfied. Note that due to the requirements of voice traffic, the minimum SNR requirements have to be satisfied at all times and no scheduling is considered. In [24], this analysis is extended to a multi-cell CDMA and TDMA/CDMA networks.

Several papers [45, 60, 64, 61, 62] discuss the implementation of these ideas to the field of multi-hop wireless networks. One of the most recent protocols defined along these lines is CA/CDMA [61]. In this protocol, every transmitter transmits with the minimum necessary power, increased by some margin. This margin allows it to resist some amount of interference caused by concurrent transmissions. The MAC protocol guarantees that the interference from concurrent transmissions is not going to exceed the margin and organizes medium access accordingly. If it happens that two nodes cannot access the medium at a same time due to power and SNR constraints, the MAC will allocate different slots for each of them. Protocol in [61] considers a joint scheduling and power adaptation. It does not consider rate adaptation. Also, routing is not considered explicitly and is left for an independent layer.

We see from the above description that the main difference between power control protocols in cellular and multi-hop networks is in scheduling. Contrary to cellular networks where all nodes constantly talk to the base station, network protocols for multi-hop network have a considerably more complex task as they need to control power and medium access at the same time.

The works presented in this section have an improved modeling of interference. Each link can have an arbitrary transmission power, and the total interference on a node is expressed as a sum of all concurrent transmissions. The success of a transmission will depend on the level of the cumulative interference. The main drawback of these models is that they do not consider rate adaptation. In the case of best effort traffic, this is an important degree of freedom that is not exploited.

Variable-Rate, Fixed-Power Protocols

Another class of protocols [81, 35, 33, 43, 50] are focused on rate adaptation: the transmission power is still kept fixed, but the rate is adapted to the actual channel conditions and the amount of interference. The rate can be adapted through variable coding or modulation, which are typically controlled at the physical layer.

The medium access control in this case is kept the same as in the original 802.11 MAC: RTS/CTS packets are sent at the beginning of each transmission and they are encoded with the lowest available rate (the most error-resistant coding and modulation). All the nodes that receive an RTS/CTS packet stay silent, and since the lowest rate is heard on a large surface, most of the interference is eliminated. Each source then adapts the transmission rate to the channel conditions on a link. The aim of this rate adaptation is thus to adapt to mobility and channel fading, and not to adapt to possible concurrent, interfering transmissions (which are already eliminated by mutual exclusion). There is no joint rate adaptation and medium access.

The protocol first enforces mutual exclusion and eliminates interference, and then it adapts rates accordingly. The same principles are also applied in the 802.11a/b/g design.

The major drawback of these protocols is that mutual exclusion and power are not adapted to network conditions. In all of the rate-adaptation based protocols, mutual exclusion is implemented through RTS/CTS packets where RTS/CTS are assumed to be sent at the lowest rate possible. This way, the total interference at a receiver is small, and the link rate is high. However, many nodes are excluded and the spatial reuse might be low.

To illustrate the problem we turn to [48], where a similar problem is discussed within the framework of HDR/CDMA cellular networks. Kumaran and Qian concluded that there exists an optimal exclusion region around a destination (base-station). The optimal strategy is to have nearby nodes with strong signals transmitting alone in a given slot, whereas distant nodes with weaker signals should be grouped and transmit during the same time slot.

Given these conclusions, it is not clear if the existing rate adaptation and mutual exclusion strategies are indeed optimal. Furthermore, packets are always sent with full power, which also decreases a potential spatial reuse, and the optimality of this decision is also not clear. Finally, in all these protocols, routing is not considered and is assumed to be implemented as a separate layer.

Variable-Rate, Variable-Power Protocols

The third group of protocols tries to adapt both rate and power [88, 70, 19, 58, 65]. Although a rate is adapted to a level of interference at a receiver, both power adaptation and medium access are used concurrently as a way to manage interference at receivers.

One of the models that incorporate rate and power adaptation is presented in [70]. Qiu and Chawla analyze the joint rate and power adaptation problem for cellular networks. They assume rates can be tuned by choosing an appropriate modulation. The main drawback is that they do not consider mutual exclusion but assume all nodes constantly access the medium with the same powers and rates. Also, since they considered a cellular setting, there is no routing problem.

One of the most general models of a wireless network that covers both routing, scheduling and power control is defined in [88] by Toumpis and Goldsmith. This model also includes different physical layer models, such as successive decoding. Toumpis and Goldsmith take the total network capacity as a performance measure. However, the complexity of the model is such that even with a linear objective function, it is difficult to solve for more than 10 nodes. It is thus difficult to draw a general conclusion about network design from such small networks.

The same authors in [89] discuss an extended model that analyzes the interaction between

the optimal routing and the optimal MAC protocol is given in in the case of CSMA/CA networks with different back-off algorithms.

In [19], Cruz and Santhanam consider a network with scheduling and a best-effort traffic. They search for the optimal scheduling, rate and power control that maximizes the total throughput of a network, given the power constraints for each user. They show that the optimal power control consists in sending with maximum power. They further solve the scheduling problem for small networks and they demonstrate a distributed algorithm that finds an approximate solution for large networks based on hierarchy. In another paper [20], they consider the same problem jointly with routing, but in the power minimization framework.

The main drawback of this model, as well as of [88], is the design objective. As discussed in Chapter 4, maximizing the total capacity is not an appropriate metric for multi-hop wireless networks design.

Proportional fairness and maximizing the minimal rate in a network (a weaker version of max-min fairness) are analyzed in [44]. However, the latter considers only a subset of possible routing and MAC protocols, those that can be transformed to convex problems. It does not apply to the most general network descriptions analyzed in other papers.

There is another class of models whose primary goal is to analyze the stability of networks. These models are very general and cover all possible network implementations. However, the models are difficult to solve, and their main goal is not to find the jointly optimal power allocation, rate allocation, scheduling and routing. Instead, they focus on analyzing the stability of a class of protocols. Examples of such models are [65] and [52].

One of such models is presented in [65]. It does not consider routing, but it considers the most general form of rate adaptation, power adaptation and scheduling. Neely, Modiano and Rohrs give stability criteria for network protocols, which are independent of a choice of a performance objective. However, they do not give an explicit solution to the power allocation and scheduling subproblems.

A general optimization framework that includes routing is given in [52]. Here, Lin and Shroff decompose the routing and the scheduling subproblems. They present the optimal routing policy and show its stability. This policy assumes the optimal scheduling and power control, which remains unknown.

In [21] Cuomo et. al. consider a UWB network without scheduling (with a single power allocation). They adapt both rate and powers, and consider both QoS and best effort traffic. They show that in the optimal power allocation, which maximizes the total throughput of a network, each node sends with full power or does not send at all.

Another protocol that considers both rate and power adaptation is DCC-MAC [58]. It is based on the results of this thesis.

2.3 Discussion

In Section 2.1 we have seen that there are three main performance objectives for network design. Furthermore, it has been observed that there is a tension between efficiency and fairness in each of these objectives. This issue is well understood in the case of wired networking, and much less in the wireless networks. One of the goals of this thesis, as explained in Chapter 1, is to understand well these trade-offs and to be able to suggest which objective should be used in wireless network design.

We have also reviewed the existing wireless protocols from the point of view of the main protocol building blocks: routing, scheduling, power control and rate control. We have seen that most of the protocols use only a few of these building blocks and assume others are fixed. Also, in many cases, even where several building blocks are used, they are designed independently and not jointly, hence the design is often suboptimal.

There are, however, existing works that formulate the most general model of a wireless network. In some cases, the model is too complicated and is not optimized (but is used, for example, to prove certain stability issues). In other cases, when the optimal design can be found, it can be done for small networks and it can use complex centralized algorithms that do not give insights into protocol design.

Our goal is, as defined in Chapter 1, to adopt the most general network model, and using this model to define cross-layer network design as an optimization problem. Analyzing the optimization problem, we want to derive guidelines for optimal protocol design. These guidelines may be suboptimal and may be derived using different heuristic. However, they should be simple such that they can be used in actual protocol design.

Chapter 3

A Unified Framework for Max-Min and Min-Max Fairness with Applications

3.1 Introduction

In the first part of the thesis, in Chapter 3 and Chapter 4, we analyze different rate-based design objectives for multi-hop wireless networks. The goal is to find an appropriate design objective that we will later use in network design. One of the frequently used design objectives is max-min fairness. However, it is not known if and in which wireless settings max-min fairness can be defined, and how it can be obtained. The goal of this chapter is to answer to these questions.

Max-min fairness is widely used in various areas of networking. In every case where it is used, there is a proof of existence and one or several algorithms for computing the max-min fair allocation; in most, but not all cases, they are based on the notion of bottlenecks. In spite of this wide applicability, there are still examples, arising in the context of wireless or peer-to-peer networks, where the existing theories do not seem to apply directly.

In this chapter, we give a unifying treatment of max-min fairness, which encompasses all existing results in a simplifying framework and extends its applicability to new examples. First, we observe that the existence of max-min fairness is actually a geometric property of the set of feasible allocations (uniqueness always holds). There exist sets in which max-min fairness does not exist, and we describe a large class of sets in which a max-min fair allocation does exist. This class contains the compact, convex sets of $\mathbb{R}^N$, but not only.

Second, we give a general purpose, centralized algorithm, called max-min Programming, for computing the max-min fair allocation in all cases where it exists (whether the set of feasible allocations is in our class or not). Its complexity is of the order of N linear programming steps

in $\mathbb{R}^N$, in the case where the feasible set is defined by linear constraints.

We show that, if the set of feasible allocations has the free-disposal property, then max-min programming degenerates to a simpler algorithm, called water-filling, whose complexity is much less. Free disposal corresponds to the cases where a bottleneck argument can be made, and water-filling is the general form of all previously known centralized algorithms for such cases. Our derivations are based on the relation between max-min fairness and leximin ordering. All our results apply mutatis mutandis to min-max fairness. Our results apply to weighted, unweighted and util-max-min and min-max fairness.

All algorithms presented here are centralized and formulated on the unified model. Applications of the algorithms to compute the max-min fair allocation on real-world networking examples, and their distributed implementations, are outside the scope of this chapter.

This chapter is organized as follows. In the next section we explain our motivation for the research presented in this chapter. We review the existing results on max-min fairness and show, by several examples, why these results cannot always be always applied.

In Section 3.3 we precisely define our framework (max-min and min-max fairness in N continuous variables). We mention a number of elementary results, such as the uniqueness and the reduction of weighted max-min fairness to the unweighted case. We recall the definition of leximin ordering that we use in a latter analysis. In this section we prove our first main result about the existence of max-min fairness. At the end of the section we also show that max-min fairness is a limiting case of utility fairness whenever a feasible set is compact.

In Section 3.4, we give the definitions of the two analyzed algorithms: max-min programming (MP) and water-filling (WF). We prove that max-min programming finds the max-min fair allocation whenever it exists, and we show that it degenerates to water-filling whenever the feasible set has the free disposal property.

In Section 3.5 we illustrate the application of our framework on three networking examples. One example is on minimum-energy wireless networks. The second one is on load distribution in P2P systems. The last example is maximum-lifetime sensor networks.

We summarize the main findings and conclude in Section 3.6.

3.2 Motivation

3.2.1 Max-min Fairness

Max-min fairness is a simple, well-recognized approach to define fairness in networks [8]; it aims at allocating as much as possible to poor users, and, at the same time, not unnecessarily wasting resources (see Section 3.3.1 for a formal definition). It was used in window flow control protocols [23], then became very popular in the context of bandwidth sharing policies for ABR service in ATM networks [17]. It is now widely used in various areas of networking [37, 13, 36, 40, 31, 78, 68, 86, 23, 15].

One of the simplest max-min fairness examples, given in [8], is single-path rate allocation. Suppose we have a network consisting of links with fixed capacities, and a set of source destination pairs that communicate over a single path each, and with fixed routing. The problem is to allocate a rate to each source-destination pair, while keeping the rate on each link below capacity. We call a rate allocation max-min fair if one cannot increase the rate of a flow without decreasing the rate of an already smaller flow (it is precisely defined in Definition 3.1). A set of feasible rate allocations for a simple two source example is given in Figure 3.1. A definition dual to a max-min fair allocation is min-max fair allocation, and is used in the context of workload distribution, where the goal is to spread a given workload evenly to all the parties: see [49], and where rates have to be allocated to available links as evenly as possible.

3.2.2 Microeconomic Approaches to Fairness

Microeconomic theories of social welfare functions and social optima discuss a fair choice of alternatives (such as goods distribution or policy making) [54]. Each possible alternative is assigned a utility, that represents its value to each individual in the system. A social welfare function is a way to aggregate individual utilities into a social utility. The optimal choice of the alternative is the one that maximizes the social welfare function [54].

There are numerous ways to define social welfare functions. One is a maximin or Rawlsian social welfare function [76] that maximizes the utility of the worst-off individual. It has been widely used in the design of communication systems (see for example [57]).

The main problem of the maximin social welfare function is that the optimal alternative is not necessarily Pareto optimal. In other words, starting from the maximin optimal alternative one can increase the utility of one individual without decreasing utilities of the others, and this is clearly not a desirable property of an efficient alternative.

A leximin social welfare ordering is a refinement of the maximin social welfare function [16, 10]. It is based on the notion of the *leximin ordering*: one vector is said to be leximin larger or equal than the other if its ordered permutation is lexicographically larger or equal to the ordered permutation of the other vector (a precise definition is given in Definition 3.4 in Section 3.3.2). The leximin social welfare optimum is always Pareto optimal [54].

The fairness criteria in networking are based on findings from social welfare theory. Max-min fairness is closely related to leximin ordering. We discuss this issue in depth in Section 3.3.2. Other concepts of fairness, like proportional fairness [47], are also based on social welfare theory. We discuss them more in Section 3.3.4.

Another important concept from microeconomics we use is the *free disposal property*. In economics, it is defined as the right of each user to dispose of an arbitrary amount of owned commodities [54], or alternatively, to consume fewer resources than maximally allowed. The formal definition is given in Definition 3.6 in Section 3.4.2.

3.2.3 Bottleneck and Water-Filling

Most of the existing works on max-min fairness rely on the notion of bottleneck link. Referring again to the single-path rate allocation example given in Figure 3.1, we say that a link is a bottleneck for a given flow if the flow uses the link, if the link is fully utilized, and if the flow has the maximal rate among all the flows that use the link. It is shown in [8] for the above example that if each flow has a bottleneck link, then the rate allocation is max-min fair. This finding, which we call the bottleneck argument, is often used to prove the existence of max-min fairness.

The most widely used algorithm for obtaining max-min fairness is the *water-filling* algorithm (WF) [8]. The principles of WF are the following: rates of all flows are increased at the same pace, until one or more links are saturated. The rates of flows passing over saturated links are then frozen, and the other flows continue to increase rates. The algorithm is repeated until all rates are frozen. A more precise description of WF algorithm is given in Section 3.4.2. It is proven in [8] that the output of WF, applied on a wired network, yields the max-min fair allocation.

A simple example of WF in 2 dimensions on a wired network with single-path routing is given in Figure 3.1. We see in the example that although WF, as defined in [8], is related to the network topology, max-min fair allocation itself is solely a property of the set of feasible rates.

An extension of this scenario is introduced for example in [40] and [36]. Each flow is separately guaranteed a minimal rate. The algorithm used there for computing the max-min fair rate allocation is a modified WF. Specifically, all rates are set to their minimal guaranteed values, and only the lowest rates are increased. A simple 2-dimensional example with an illustration of WF

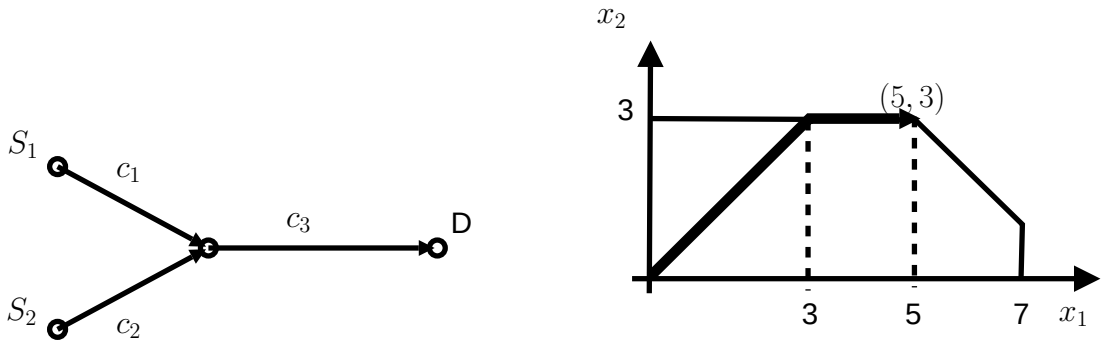


Figure 3.1: An example of a network with its feasible rate set, and max-min allocation obtained by water-filling. On the left, a network of 3 links is given. Flow x_1 connects S_1 and D and passes over links 1 and 3, and flow x_2 connects S_2 and D and traverses links 2 and 3. The set of feasible rates (x_1, x_2) is depicted by the inequalities: $0 \leq x_1 \leq c_1$, $0 \leq x_2 \leq c_2$ and $x_1 + x_2 \leq c_3$, and it is given on the right ($c_1 = 7, c_2 = 3, c_3 = 8$). The water-filling, as explained in [8], is depicted by the bold arrow: first, rates on both flows are increased, until flow x_2 hits the limit $x_2 = 3$. Then, x_2 is frozen, and x_1 is further increased until we reach the equality $x_1 + x_2 = 8$. The max-min fair rate allocation is $(5, 3)$.

is given on the left of Figure 3.2.

Max-min fairness for single-rate multicast sessions is defined in [31]. This is generalized to multi-rate multicast sessions in [78]. Rates are again upper-bounded by links' capacities, and here we are interested in max-min fair allocation of receivers rates. A set of feasible allocations is linearly constrained, and a WF approach can be used. The geometric shape of the feasible set is essentially the same as in single-path routing.

What the forementioned scenarios have in common is the linearity of the constraints defining the feasible set. In [13], a single-path routing scenario is considered, and each source is assigned a utility, which is an increasing and concave function of its rate. Instead of searching for a max-min fair rate allocation, they look for max-min fair utility allocation. This approach is generalized in [78], where a max-min fair utility allocation is considered in the context of a multicast network. Here, the authors only required that a utility function be a strictly increasing but not necessarily concave function of rate, hence the feasible set is not necessarily convex. A simple 2-dimensional example is given in Figure 3.2, on the right. We also see that the WF algorithm can be used in this case as well.

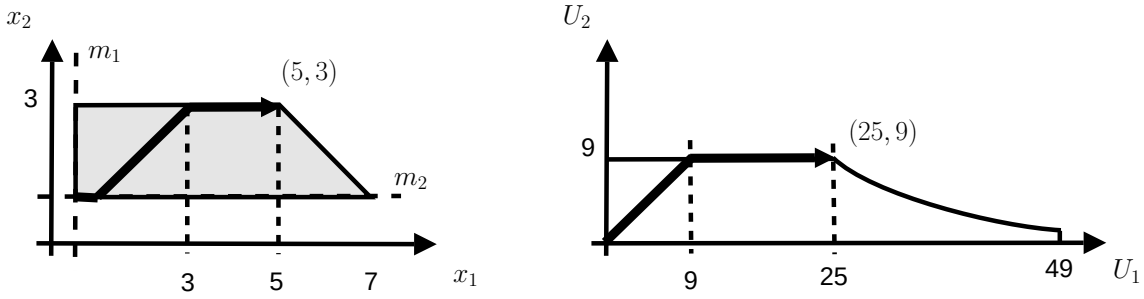


Figure 3.2: More examples of feasible rate sets. We consider again the topology given on the right of Figure 3.1. We first assume there are minimum rates guaranteed, $m_1 = 0.5$ and $m_2 = 1$, for flows x_1 and x_2 respectively. The feasible set for this case is depicted on the left. The water-filling, as explained in [40, 36], is represented with the bold arrow: we start by setting both flows on their minimal values ($x_1 = m_1, x_2 = m_2$). We start by increasing the smaller rate, x_1, until we reach $x_1 = x_2 = 1$. The rest is the same as in the example in Figure 3.1. On the right we consider utility max-min fairness, presented in [13, 78]. Again, network is the same given in Figure 3.1, and we assume utility function is $U(x) = x^2$. The set of feasible utilities is depicted on the right. It is a non-convex set. The max-min fair utility allocation can be obtained by water-filling, similarly as in Figure 3.1: utilities on both flows are increased, until flow x_2 hits the limit $x_2 = 3$. Then, x_2 is frozen, and x_1 is further increased until we reach the equality $x_1 + x_2 = 8$. The max-min fair utility allocation is $(25, 9)$ which corresponds to the rate allocation $(5, 3)$.

3.2.4 When Bottleneck and Water-Filling Become Less Obvious

It is not always obvious how to generalize the notion of a bottleneck link and the water-filling approach to an arbitrary problem. To see why, consider a point-to-point multi-path routing scenario, where, to our knowledge, max-min fairness was not studied before. We look at the same set-up as above, but now allow for multiple paths to be used by a single source-destination pair. The rate of communication between a source and a destination is equal to the sum of rates over all used paths, and we still want a max-min fair rate allocation of source-destination rates. This is defined as above, in other words, a source-destination rate allocation is max-min fair if one cannot increase the rate of a source-destination pair without decreasing the rate of some other pair, which is already smaller. An example is given in Figure 3.3. Let us assume that all links in the given example have capacity 1. Then, by increasing all the rates at the same pace, we will have rates of all paths equal to 1/2 when link 3-4 saturates. Now, if we continue increasing the rate over path 1-4, the rate of source destination pair 1 will be higher than the rate of source destination 2, and path 2-3-4 will loose its bottleneck since it is no longer the biggest end-to-end flow that uses 3-4. If we change the definition of the bottleneck, and instead of taking the biggest

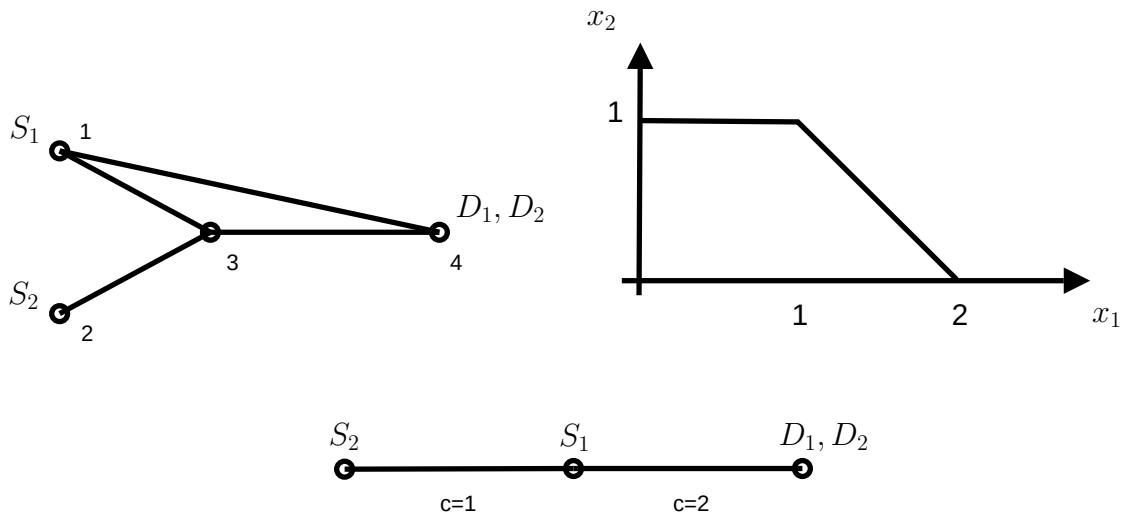


Figure 3.3: A simple multi-path example. Up, left: S1 sends to D1 over two paths, 1-3-4 and 1-4, while S2 sends to D2 over a single path 2-3-4. Up, right: the set of feasible rates. Down, center: the corresponding virtual single path problem.

end-to-end flow, we consider the path with the highest rate, we obtain the max-min fair path rate allocation that differs from the end-to-end max-min fair rate allocation.

A question that arises is how to define a bottleneck in this case, such that the water-filling algorithm still works, if it is possible at all. Also, it is not clear if for a given definition of a bottleneck we can still claim the same as for the single-path case, i.e., that if each path has a bottleneck, then the allocation is max-min fair. Finally, we do not even know, using the existing state of the art, if a max-min fair allocation exists on an arbitrary multi-path network.

This example can be solved by observing that the max-min fair allocation depends only on the set of feasible rates. Consider again the example in Figure 3.3, left. Assume all links have unit capacity. Call $x_1 = y_1 + y_2$ the rate of source 1, and x_2 the rate of source 2, where y_1 is the rate of source 1 on path $1-4$, and y_2 on path $1-3-4$. The set of feasible rates is the set of $(x_1 \geq 0, x_2 \geq 0)$ such that there exist slack variables $y_1 \geq 0, y_2 \geq 0$ with $y_1 \leq 1, y_2 + x_2 \leq 1$ and $x_1 = y_1 + y_2$. This is an *implicit* definition, which can be made explicit by eliminating the slack variables; this gives the conditions $x_1 \leq 1, x_1 + x_2 \leq 2$ (Figure 3.3, center). The set is convex, with a linear boundary, as in Figure 3.1, left. We can re-interpret the original multi-path problem as a virtual single path problem (Figure 3.3, right), and apply the existing WF algorithms. Note however that the concept of bottleneck in the virtual single path problem has lost its physical interpretation.

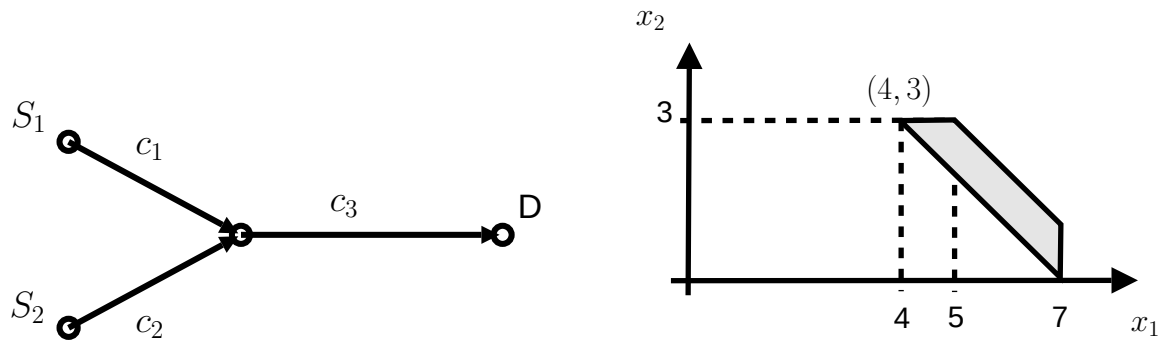


Figure 3.4: When water-filling does not work - consider the network topology on the left ($c_1 = 7, c_2 = 3, c_3 = 8$). Suppose that node D receives parts of the same stream from both S_1 and S_2, through flows x_1 and x_2, and suppose it needs a minimal total rate of $x_1 + x_2 \geq 7$. We want to minimize loads of servers S_1 and S_2, and we are interested in min-max fair allocation of (x_1, x_2). The feasible rates set is given on the right. Min-max fair allocation exists, and it is $(4, 3)$.

3.2.5 When Bottleneck and Water-Filling Do Not Work

Unfortunately, the trick applied when modifying the multi-path problem into a single path problem does not always work. Consider the following workload distribution example, where min-max fairness is used: servers in a peer-to-peer network send data to a client; every client receives data from multiple servers, and has a guaranteed minimal rate of reception. Each flow from a server to a client is constrained by link capacities. We are here in a best effort scenario, like in traditional networking cases, but here users are providing resources (the servers), and a natural definition of fairness in this setting is min-max fairness, where we try to give the least possible work to the most loaded server. We are thus searching for a min-max distribution of server load. A 2-dimensional example is given in Figure 3.4. A simple visual attempt to apply WF on the set of feasible rates shows that it is not possible. We discuss this example in more detail in Section 3.4.2 and Section 3.5.2.

A similar, but simpler, example is given in [49], which focused on finding a leximax minimal allocation (we show in Section 3.4 that the leximax minimal allocation obtained in [49] is in fact min-max fair). Its complexity is of the order of N polynomial steps in $\mathbb{R}^N$, in the case where the feasible set is defined by linear constraints.

In Section 3.5.3 we present another example where water-filling does not work. We consider the lifetime of nodes in a sensor network, inspired by the example introduced in [14], which studied the minimum lifetime. We study instead the max-min fair allocation of lifetimes. We show that the set of possible lifetimes of nodes, and that the max-min fair lifetime allocation

exists. However, as we also show, it is not possible to obtain it by water-filling.

3.3 Max-Min and Min-Max Fairness in Euclidean Spaces

3.3.1 Definitions and Uniqueness

Consider a set $\mathcal{X} \subset \mathbb{R}^N$. We define the max-min and min-max fair vectors with respect to set $\mathcal{X}$ as follows:

Definition 3.1 *[8] A vector* $\mathbf{x} \in \mathcal{X}$ *is "max-min fair on set* $\mathcal{X}$*" if and only if*

$$(\forall \mathbf{y} \in \mathcal{X})(\exists s \in \{1, \cdots, N\})\ y_s > x_s \quad \Rightarrow \quad (\exists t \in \{1, \cdots, N\})\ y_t < x_t \leq x_s \tag{3.1}$$

i.e. increasing some component x_s *must be at the expense of decreasing some already smaller or equal component* x_t*.*

Definition 3.2 *A vector* $\mathbf{x}$ *is "min-max fair on set* $\mathcal{X}$*" if and only if*

$$(\forall \mathbf{y} \in \mathcal{X})(\exists s \in \{1, \cdots, N\})\ y_s < x_s \quad \Rightarrow \quad (\exists t \in \{1, \cdots, N\})\ y_t > x_t \geq x_s \tag{3.2}$$

i.e. decreasing some component x_s *must be at the expense of increasing some already larger component* x_t*.*

It is clear that if $\mathbf{x}$ is a min-max fair vector on $\mathcal{X}$, then $-\mathbf{x}$ is max-min fair on $-\mathcal{X}$ and vice versa. Thus, in the remainder of the paper, we give theoretical results only for max-min fairness. Uniqueness is assured by the following proposition, the proof of which is analogous to the one in [8] and is not given here.

Proposition 3.1 *If a max-min fair vector exists on a set* $\mathcal{X}$*, then it is unique.*

Weighted min-max fairness is a classical variation, defined as follows. Given some positive constants w_i (called the "weights"), a vector $\mathbf{x}$ is "weighted-max-min fair" on set $\mathcal{X}$, if and only if increasing one component x_s must be at the expense of decreasing some other component x_t such that $x_t/w_t \leq x_s/w_s$ [8]. This is generalized in [78], which introduces the concept of "util max-min fairness": given N increasing functions $\phi_i : \mathbb{R} \to \mathbb{R}$, interpreted as utility functions, a vector $\mathbf{x}$ is "util-max-min fair" on set $\mathcal{X}$ if and only if increasing one component x_s must be at the expense of decreasing some other component x_t such that $\phi_t(x_t) \leq \phi_s(x_s)$ (this is also called

"weighted max-min fairness" in [68]). Consider the mapping ϕ defined by

$$(x_1, \cdots, x_N) \rightarrow (\phi_1(x_1), \cdots, \phi_N(x_N)) \tag{3.3}$$

It follows immediately that a vector $\mathbf{x}$ is util-max-min fair on set $\mathcal{X}$ if and only if $\phi(\mathbf{x})$ is max-min fair on the set $\phi(\mathcal{X})$, the case of weighted max-min fairness corresponding to $\phi_i(x_i) = x_i/w_i$. Thus, we now restrict our attention to unweighted max-min fairness.

3.3.2 Max-Min Fairness and Leximin Ordering

In the rest of our paper we will extensively use leximin ordering, a concept we borrow from economy, and which we now recall. Let us define the "order mapping" $\mathcal{T} : \mathbb{R}^N \rightarrow \mathbb{R}^N$ as the mapping that sorts $\mathbf{x}$ in non-decreasing order, that is: $\mathcal{T}(x_1, \cdots, x_n) = (x_{(1)}, \cdots, x_{(n)})$, where $(x_{(1)}, \cdots, x_{(n)})$ is a permutation of $(x_1, \cdots, x_n)$ such that $x_{(1)} \leq x_{(2)} \cdots \leq x_{(n)}$. Let us also define the lexicographic ordering of vectors in $\mathcal{X}$ by $\mathbf{x} \overset{lex}{>} \mathbf{y}$ if and only if $(\exists i)\, x_i > y_i \text{ and } (\forall j < i)\, x_j = y_j$. We also say that $\mathbf{x} \overset{lex}{\geq} \mathbf{y}$ if and only if $\mathbf{x} \overset{lex}{>} \mathbf{y}$ or $\mathbf{x} = \mathbf{y}$. This latter relation is a total order on $\mathbb{R}^N$.

Definition 3.3 *[54] Vector* $\mathbf{x}$ *is* leximin larger *than or equal to* $\mathbf{y}$ *if* $\mathcal{T}(\mathbf{x}) \overset{lex}{\geq} \mathcal{T}(\mathbf{y})$.

Definition 3.4 *[54] Vector* $\mathbf{x} \in \mathcal{X}$ *is* leximin maximal *on a set* $\mathcal{X}$ *if for all* $\mathbf{y} \in \mathcal{X}$ *we have* $\mathcal{T}(\mathbf{x}) \overset{lex}{\geq} \mathcal{T}(\mathbf{y})$.

Note that a leximin maximum is not necessarily unique. See Figure 3.5 on the left for an example.

Proposition 3.2 *[82] Any compact subset of* $\mathbb{R}^n$ *has a leximin maximal vector.*

It has been observed in [13, 40, 78] that a max-min fair allocation is also leximin maximal, for the feasible sets defined in these papers. It is generalized to an arbitrary feasible set in [82], as follows.

Proposition 3.3 *If a max-min fair vector exists on a set* $\mathcal{X}$*, then it is the unique leximin maximal vector on* $\mathcal{X}$*.*

Thus, the existence of a max-min fair vector implies the uniqueness of a leximin maximum. The converse is not true: see Figure 3.5, right, for an example of a set with unique leximin maximal vector which is not max-min achievable. [82] defines a weaker version of max-min fairness, "maximal fairness"; it corresponds to the notion of leximin maximal vector, hence it is

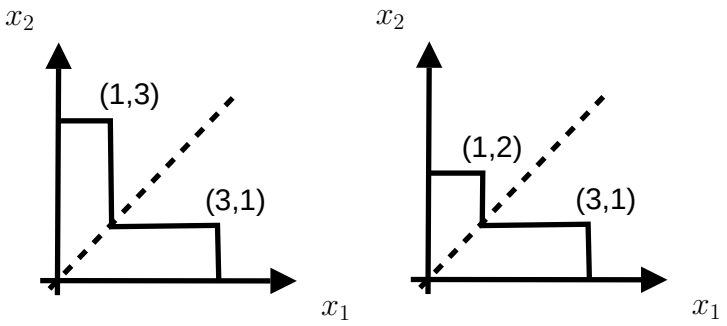


Figure 3.5: Examples of 2-dimensional sets that do not have max-min fair allocation. Point $(1,3)$ is not max-min fair in the example on the left since there exists point $(3,1)$ that contradicts with definition Definition 3.1. Both points $(1,3)$ and $(3,1)$ are leximin maximal in this example. In the example on the right, point points $(3,1)$ is the single leximin maximal point. Still, it is not the max-min fair point. Note that there exist no real networking example we are aware of that has these feasible rate sets – these sets are only artificial examples that illustrate properties of leximin ordering.

not unique, and exists on a larger class of feasible sets. We leave this weaker version outside the scope of this paper.

It is shown in [54] that if a vector is leximin maximal, it is also Pareto optimal. Therefore, from Proposition 3.3 it follows that max-min fair vector, if it exists, is Pareto optimal. The converse is not necessarily true.

3.3.3 Existence and Max-Min Achievable Sets

As already mentioned, a number of papers showed the existence of max-min fair allocation in many cases, using different methods. We give here a generalized proof that holds on a larger class of continuous sets. This class of continuous sets includes the feasible sets of all the networking applications we are aware of. Note that a max-min fair vector does not exist on all feasible sets, even compact and connected. Simple counter-examples are given in Figure 3.5. However, these counter-examples are artificial and do not correspond to any networking scenario. In the reminder of this section we give a sufficient condition for a max-min vector to exist.

Definition 3.5 *A set $\mathcal{X}$ is max-min achievable if there exists a max-min fair vector on $\mathcal{X}$.*

Theorem 3.1 *Consider a mapping ϕ defined as in Equation (3.3). Assume that ϕ_i is increasing and continuous for all i. If the set $\mathcal{X}$ is convex and compact, then $\phi(\mathcal{X})$ is max-min achievable.*

We first give an intuition on how we shall prove the theorem. We observe vector $\mathbf{x}$ that is leximin maximal on the set $\phi(\mathcal{X})$, and we want to prove that this is at the same time the max-min fair vector. The idea of the proof is to assume the contrary, that there exists a vector $\mathbf{y}$ that violates the definition of max-min fairness of vector $\mathbf{x}$. We will then construct vector $\mathbf{z}$ from $\mathbf{x}$ and $\mathbf{y}$ such that $\mathbf{z}$ is leximin-larger than $\mathbf{x}$, which will lead to contradiction. Function $\phi()$ is strictly increasing, hence there exists and inverse $\phi^{-1}()$, which is also strictly increasing. Although set $\phi(\mathcal{X})$ is not convex, set $\mathcal{X}$ is convex. Therefore, we will chose α such that vector $\mathbf{z}$, constructed as $\phi^{-1}(\mathbf{z}) = \alpha\phi^{-1}(\mathbf{x}) + (1-\alpha)\phi^{-1}(\mathbf{y})$, is leximin larger than $\mathbf{x}$.

Proof: Let $\mathbf{x} \in \phi(\mathcal{X})$ be a vector such that for all $\mathbf{y} \in \phi(\mathcal{X})$ we have $\mathcal{T}(\mathbf{x}) \overset{lex}{\geq} \mathcal{T}(\mathbf{y})$. Such a vector exists according to proposition 3.2, since set $\mathcal{X}$ is compact. In order to prove the theorem, we proceed by contradiction, assuming that there exist $\mathbf{y}$ and an index $s \in \{1, \cdots, N\}$ such that $y_s > x_s$ and for all $t \in \{1, \cdots, N\}$, $x_t \leq x_s$ we have $y_t \geq x_t$. We then define a permutation $\pi : \{1, \cdots, N\} \rightarrow \{1, \cdots, N\}$ such that for all $i < j$, $\mathbf{x}_{\pi(i)} \leq \mathbf{x}_{\pi(j)}$, and either $x_{\pi(l)} < x_{\pi(l+1)}$ or $l = N$, where $l = \pi^{-1}(s)$. The last part of the requirement is important if there are several components of the vector that are equal to x_s, hence there are several permutations that maintain non-decreasing ordering. We then want s to be mapped by π to the largest such index: if $l = \pi^{-1}(s)$ than either $x_s < x_{\pi(l+1)}$ or l is the last index ($l = N$).

Next, let us define vector $\mathbf{z}(\alpha) = \phi(\alpha\phi^{-1}(\mathbf{x}) + (1-\alpha)\phi^{-1}(\mathbf{y}))$. Although we cannot make a convex combination of $\mathbf{x}$ and $\mathbf{y}$ since set $\phi(\mathcal{X})$ is not convex, we can make a convex combination of $\phi^{-1}(\mathbf{x})$ and $\phi^{-1}(\mathbf{y})$ in the set $\mathcal{X}$ which is convex.

It is easy to verify that for $\alpha \in (0,1)$, $\mathbf{z}(\alpha)$ belongs to $\phi(\mathcal{X})$ and for all $i \in \{1, \cdots, N\}$, $\min(x_i, y_i) < \mathbf{z}(\alpha)_i < \max(x_i, y_i)$, due to convexity of $\mathcal{X}$ and strictly increasing properties of functions ϕ_i. Also, for all i let us pick an arbitrary α_i satisfying

$$\alpha_i \in \begin{cases} \left(\frac{\phi_i^{-1}(x_s)-\phi_i^{-1}(y_i)}{\phi_i^{-1}(x_i)-\phi_i^{-1}(y_i)}, 1\right), & x_s \in [y_i, x_i), \\ [0,1), & \text{otherwise} \end{cases}$$

and we call $\alpha_m = \max_i(\alpha_i)$ and $\mathbf{z} = \mathbf{z}(\alpha_m) \in \phi(\mathcal{X})$ (since $\alpha_m \in [0,1)$). Intuitively, if for some i, $y_i < x_s$, we want to have $z_i > x_s$. If $x_i \leq x_s$ (including when $i = s$) we than by assumption have $y_i \geq x_i$, and we choose α such that we get $z_i > x_i$. Finally, if both $y_i > x_s, x_i > x_s$, than we can select any α and we will have $z_i > x_s$. A choice of constant is depicted in Figure 3.6.

We have chosen the highest of α_i, hence we now have that if $x_i \leq x_s$, than $z_i \geq x_i$, otherwise

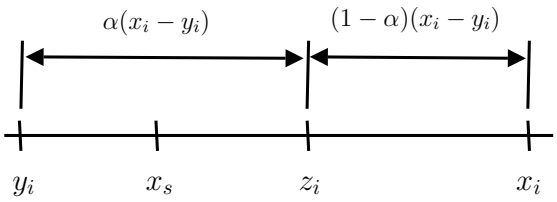


Figure 3.6: Choice of constants: suppose for simplicity that $\phi(x) = x$. We choose α such that: if for some i, $x_i > x_s$, then $z_i > x_s$ (as depicted on the figure), if $x_i \leq x_s$, then $y_i \geq z_i \geq x_i$, or else for any α, $z_i > x_s$.

$z_i \geq x_s$. We also have $z_s > x_s$. From this, we derive the following property of the sorted vectors

$$\begin{aligned} z_{\pi(i)} &\geq x_{\pi(i)}, \quad \text{for } i < l, \\ z_{\pi(i)} &> x_{\pi(l)}, \quad \text{for } i \geq l. \end{aligned}$$

We first notice that for all i, $z_{\pi(i)} \geq x_{\pi(1)}$, and as $\mathcal{T}(\mathbf{x}) \overset{lex}{\geq} \mathcal{T}(\mathbf{z})$ we conclude that $z_{(1)} = z_{\pi(1)} = x_{\pi(1)}$. Next, assuming that for some $i < l$ and for all $j < i$ we have $z_{(j)} = z_{\pi(j)} = x_{\pi(j)}$, then again as for all $j \geq i$, $z_{\pi(j)} \geq x_{\pi(i)}$, and $\mathcal{T}(\mathbf{x}) \overset{lex}{\geq} \mathcal{T}(\mathbf{z})$ we conclude that $z_{(i)} = z_{\pi(i)} = x_{\pi(i)}$. Hence, by induction we have proved that for all $i < l$ we have $z_{(i)} = z_{\pi(i)} = x_{\pi(i)}$. Finally, since for all $i \geq l$ we have $z_{\pi(i)} > x_{\pi(l)}$, hence $z_{(i)} > x_{\pi(l)}$ we necessarily have that $\mathcal{T}(\mathbf{z}) \overset{lex}{>} \mathcal{T}(\mathbf{x})$, which brings us to the contradiction.

Therefore, we conclude that a leximin maximal vector on a set $\mathcal{X}$ is also a max-min fair vector, and set $\mathcal{X}$ is max-min achievable.

q.e.d.

As a special case, obtained by letting $\phi_i(x) = x$, we conclude that all convex and compact sets are max-min achievable. Taking $\phi_i(x) = x/w_i$, we also conclude that weighted max-min fairness exists on all compact, convex sets. More generally, util-max-min fairness exists on all compact, convex sets, if the utility functions are continuous (and increasing).

Note that if $\mathbf{x}^*$ is max-min fair on $\mathcal{X}$ and $\mathbf{y}^*$ is max-min fair on $\phi(\mathcal{X})$, then in general $\mathbf{y}^* \neq \phi(\mathbf{x}^*)$. To see this, consider an example of a wired network with a single link of capacity c, with two flows, x_1 and x_2 passing over the link. The max-min fair rate allocation in this case is $x_1 = x_2 = c/2$. Let us next define the utility functions $\phi_i(x) = x/w_i$, as above, and let $w_1 = 1, w_2 = 2$. We then have $u_1 = x_1, u_2 = x_2/2$, and the max-min fair utility allocation is going to be $u_1 = u_2 = c/3$ and the resulting rate allocation is going to be $x_1 = c/3, x_2 = 2c/3$,

which differs from the max-min fair rate allocation in the first case.

In [13], the utility functions ϕ_i are arbitrary, continuous, increasing and concave functions. With these assumptions, the set $\phi(\mathcal{X})$ is also convex and compact. Note that in general, though, the set $\phi(\mathcal{X})$ used in Theorem 3.1 is not necessarily convex. Examples with non-convex sets are provided in [68] and [78].

3.3.4 Max-min Fairness as A Limiting Case of Utility Fairness

As explained in Section 3.2.2, utility fairness is an another popular class of fairness objectives in networking, derived from social welfare theory. If $\mathcal{X} \subset \mathbb{R}^N$ is the set of feasible rate allocation, and $U_i : \mathbb{R} \rightarrow \mathbb{R}$ for all $i \in \{1 \cdots N\}$, then utility fair rate allocation is the one that maximizes $\sum_{i=1}^{N} U_i(x_i)$ over $x \in \mathcal{X}$. The most well-known example of utility fairness is the proportional fairness [47], where $U_i(x) = \log(x)$. There exist many distributed algorithms for solving the convex optimization problem of finding the utility fair rate allocation [59, 53].

It is known that no continuous utility function exists that can represent leximin ordering [54], hence we cannot directly use the existing utility-based algorithms to find a max-min fair vector. However, it is shown in [59] that for wired networks with single-path routing, and for a class of utility functions defined as

$$U_\xi(x) = \begin{cases} (1-\xi)^{-1} x^{1-\xi} & \text{if } \xi \neq 1 \\ \log(x) & \text{if } \xi = 1 \end{cases}, \tag{3.4}$$

the utility fair rate allocation converges to max-min fair one as ξ goes to infinity. Therefore, by choosing an appropriate ξ, we can approach arbitrarily close to the max-min fair allocation using the existing algorithms [59, 53].

The above property of max-min fair allocation is shown to hold, in [59], only in the case of wired networks with single-path routing, and the proof of it heavily depends on the notion of bottleneck. It is not known if the same result holds for an arbitrary problem where max-min fair allocation applies. We show here that it holds whenever the feasible allocation set is convex.

Let U_ξ be a family of increasing, concave and differentiable functions defined on $\mathbb{R}^+$. Assume that, for any fixed numbers x and δ,

$$\lim_{\xi \rightarrow +\infty} \frac{U_\xi(x)}{U_\xi(x+\delta)} = 0, \tag{3.5}$$

$$\lim_{\xi \rightarrow +\infty} \frac{U'_\xi(x)}{U'_\xi(x+\delta)} = 0. \tag{3.6}$$

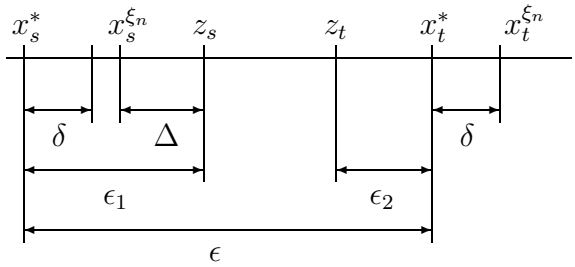


Figure 3.7: Choice of constants

The assumptions Equation (3.5) and Equation (3.6) are satisfied if for example U_ξ is defined as in Equation (3.4).

Let us consider a convex set $\mathcal{X}$ and let $\mathbf{x}^\xi$ be the utility-fair vector on $\mathcal{X}$, that is the unique vector that maximizes $\sum_{i=1}^N U_\xi(x_i)$ The following theorem shows that a max-min fair vector is then a limiting case of utility fair vector for the set of utilities we constructed above:

Theorem 3.2 *The set of utility-fair vectors $\mathbf{x}^\xi$ converges toward the max-min fair vector as ξ tends to $+\infty$.*

Proof: We first prove that if $\mathbf{x}^*$ is an accumulation point for the set of vectors $\mathbf{x}^\xi$, then $\mathbf{x}^*$ is the max-min fair vector, and to prove that we assume the contrary, that exists a vector $\mathbf{y}$ and a corresponding index $s = s(y)$ such that $y_s > x_s$ and for all $t \in S$ $x_t \leq x_s \Rightarrow x_t \leq y_t$. Let us call $T_s(x) = \{t \in \{1, \cdots, N\} \,|\, x_t > x_s\}$. If set $T_s(x)$ is empty, we pick an arbitrary $\epsilon > 0$, else we call $t = \arg\min_{t \in T_s(x)} x_t^*$ and $\epsilon = x_t^* - x_s^*$.

Next, similarly as in proof of theorem 3.1, we pick an arbitrary $\alpha < \epsilon/(y_s - x_s^* + x_t^* - y_t)$, and we call $\mathbf{z} = (1-\alpha)\mathbf{x} + \alpha\mathbf{y}$. If $\epsilon_1 = z_s - x_s^*$ and $\epsilon_2 = x_t^* - z_t$, it is easy to verify that $\epsilon > \epsilon_1 + \epsilon_2$. Also, $\mathbf{z} \in \mathcal{X}$ due to convexity of $\mathcal{X}$.

Finally we pick an arbitrary $\delta > 0$, $\Delta > 0$ such that $\delta + \Delta < \epsilon_1$, and some n_0 such that for all $n \geq n_0$ and for all $u \in \{1, \cdots, N\}$ we have $|U_{\xi_n}(x_u^*) - U_{\xi_n}(x_u^{\xi_n})| < \delta$. The choice of the above constants is depicted in Figure 3.7.

Now we show that for n large enough we have a contradiction with optimality of $\mathbf{x}^{\xi_n}$. Consider the expression A defined by

$$A = \sum_{t \in \{1, \cdots, N\}} \left(U_{\xi_n}(z_t) - U_{\xi_n}(x_t^{\xi_n}) \right).$$

From the optimality of $\mathbf{x}^{\xi_n}$ we have $A \leq 0$. Also

$$A \;\geq\; U_{\xi_n}(z_s) - U_{\xi_n}(x_s^{\xi_n}) + \sum_{t \in T_s(x)} \left(U_{\xi_n}(z_t) - U_{\xi_n}(x_t^{\xi_n}) \right). \tag{3.7}$$

From the theorem of intermediate values, there exist numbers c_s^n and c_t^n for all $t \in T_s(x)$ such that

$$\begin{aligned} & x_s^{\xi_n} \leq c_s^n \leq z_s, \\ & U_{\xi_n}(z_s) - U_{\xi_n}(x_s^{\xi_n}) = U'_{\xi_n}(c_s^n)(z_s - x_s^{\xi_n}). \end{aligned}$$

and

$$\begin{aligned} & z_t \leq c_t^n \leq x_t^{\xi_n}, \\ & U_{\xi_n}(x_t^{\xi_n}) - U_{\xi_n}(z_t) = U'_{\xi_n}(c_t^n)(x_t^{\xi_n} - z_t). \end{aligned}$$

where U'_{ξ_n} is the right derivative of U_{ξ_n}. From the above we have

$$\begin{aligned} U_{\xi_n}(z_s) - U_{\xi_n}(x_s^{\xi_n}) &> \Delta U'_{\xi_n}(z_s), \\ U_{\xi_n}(x_t^{\xi_n}) - U_{\xi_n}(z_t) &< (\delta + \epsilon_2) U'_{\xi_n}(z_t). \end{aligned}$$

thus from eq. (3.7)

$$\begin{aligned} A &> \Delta U'_{\xi_n}(z_s) - (\delta + \epsilon_2) M U'_{\xi_n}(z_t) \\ &> U'_{\xi_n}(z_s) \left(\Delta - (\delta + \epsilon_2) M \frac{U'_{\xi_n}(z_t)}{U'_{\xi_n}(z_s)} \right) \end{aligned}$$

where M is the cardinality of set $T_s(x)$. Now from eq. (3.5), the last term in the above equation tends to Δ as n tends to infinity. Now $U'_{\xi_n} > 0$ from our assumption thus, for n large enough, we have $A > 0$, which is the required contradiction.

The set of vectors $\mathbf{x}^\xi$ is in a compact set $\mathcal{X}$ hence it has at least once accumulation point. Since each accumulation point is also max-min fair, and from property 3.1 we have the uniqueness of a max-min fair vector we conclude that the set of vectors $\mathbf{x}^\xi$ has a unique accumulation point hence

$$\lim_{\xi \to +\infty} \mathbf{x}^\xi = \mathbf{x}^*.$$

q.e.d.

3.4 Max-Min Programming and Water-Filling

3.4.1 The Max-Min Programming (MP) Algorithm

We now present the max-min programming (MP) algorithm, which finds the max-min fair vector on any feasible set, if it exists.

1. let $S^0 = \{1, \cdots, N\}, \mathcal{X}^0 = \mathcal{X}, \mathcal{R}^0 = \mathcal{X}$ and $n = 0$
2. **do**
3. $n = n + 1$
4. Problem MP^n: maximize T^n subject to:
$$\begin{cases} (\forall i \in S^{n-1}) & x_i \geq T^n \\ \text{for some} & \mathbf{x} \in \mathcal{X}^{n-1} \end{cases} \tag{3.8}$$
5. let $\mathcal{X}^n = \{\mathbf{x} \in \mathcal{X}^{n-1} \mid (\forall i \in S^{n-1})\, x_i \geq T^n, (\exists i \in S^{n-1})\, x_i > T^n\}$,
$\mathcal{R}^n = \{\mathbf{x} \in \mathcal{X}^{n-1} \mid (\forall i \in S^{n-1})\, x_i \geq T^n\}$
and $S^n = \{i \in \{1, \cdots, N\} \mid (\forall \mathbf{x} \in \mathcal{X}^n)\, x_i > T^n\}$
6. **until** $S^n = \emptyset$
7. return the only element in $\mathcal{R}^n$

Intuitively, the algorithm maximizes in each step the minimal coordinate of the feasible vector, until all coordinates are processed. The n-th step of the algorithm is a minimization problem, called MP^n: $\mathcal{X}^n$ represents the remaining search space, S^n represents the direction of search, in terms of coordinates that can be further increased, and $\mathcal{R}^n$ is the set that will at the end contain a single rate allocation, the max-min fair one.

Proof of Correctness

The algorithm always terminates if $\mathcal{X}$ is max-min achievable, and then $\mathcal{X}^n$ is reduced to one single element, which is the required max-min fair vector, as is proved in the following theorem:

Theorem 3.3 *The above algorithm terminates and finds the max-min fair vector on $\mathcal{X}$ if it exists, in at most N steps.*

The idea of the proof is the following. We first want to show that in every step we decrease the size of S^n, that is $S^n \subset S^{n-1}$. From this we will conclude that the algorithm finishes in at most N steps. We then show that what remains in the set R^n once the algorithm stops (that is $S^n = \emptyset$), is the max-min fair allocation.

We start by introducing two lemmas before proving the theorem.

Lemma 3.1 *For all n, $\mathbf{x}, \mathbf{y} \in \mathcal{X}^n$ and $t \in \{1, \cdots, N\}$ such that $x_t \leq T^n$, we have $y_t \geq x_t$.*

Proof: We prove lemma by induction over n. If $n = 1$, we have for all t, $x_t \geq T^1$ and $y_t \geq T^1$, hence for $x_t = T^1$, we have $y_t \geq x_t$. Next assume the above is true for $n-1$. Suppose $x_t < T^n$. We then also have $x_t \leq T^{n-1}$, hence by the induction assumption we have $y_t \geq x_t$. Finally, if for some t, $x_t = T^n$ then $y_t \geq T^n$ or else we have a contradiction with the definition of T^n.

q.e.d.

Lemma 3.2 *If $\mathbf{x}$ is max-min fair vector on $\mathcal{X}$ then for all n such that $\mathcal{X}^n \neq \emptyset$, $\mathbf{x} \in \mathcal{X}^n$. The same holds for $\mathcal{R}^n$.*

Proof: We prove lemma by induction. If $\mathbf{x} \notin \mathcal{X}^1$ then $\mathbf{x}$ is not leximin maximal, hence the contradiction. Let us next assume $\mathbf{x} \in \mathcal{X}^{n-1}$ and $\mathbf{x} \notin \mathcal{X}^n$, where $\mathcal{X}^n \neq \emptyset$. Then there exists $\mathbf{y} \in \mathcal{X}^n$ and $s \in S^n$ such that $y_s > x_s$. Also, by lemma 3.1, for all $t \in \{1, \cdots, N\}$ such that $x_t \leq T^n$, we have $y_t \geq x_t$. This contradicts the assumption that $\mathbf{x}$ is max-min fair which proves the lemma. Since $\mathcal{X}^n \subseteq \mathcal{R}^n$, we have the second claim.

q.e.d.

Proof of theorem 3.3: If set $\mathcal{X}$ is max-min achievable, it is then also bounded hence there exists a solution to each MP^i. Let us call $\mathbf{x}$ max-min fair vector on $\mathcal{X}$.

We first prove that for every n we have $S^n \subset S^{n-1}$. From (3.8) we have that

$$T^n = \min_{\mathbf{x} \in \mathcal{X}^{n-1}} \min_{i \in S^{n-1}} x_i.$$

Both minima are reached since $\mathcal{X}^{n-1}$ is compact and S^{n-1} is finite. Therefore, there exist $i^* \in S^{n-1}, \mathbf{x}^* \in \mathcal{X}^{n-1}$ such that $x^*_{i^*} = T^n$, and we have $i^* \notin S^n$, thus we proved $S^n \subset S^{n-1}$. We conclude that sequence $|S^n|$ decreases and we will have $S^n = \emptyset$ in at most N steps.

We also notice that for every $i \in \{1, \cdots, N\}$ there exists m such that $i \in S^{m-1}$ and $i \notin S^m$. From $i \in S^{m-1}$ we have that for all $\mathbf{x} \in \mathcal{X}^m$, $x_i \leq T^m$. From $i \in S^m$ we have that for all $x \in \mathcal{X}^{m-1}$ we have $x_i \geq T^m$ in the constraints for MP^m. Now as for all n, $\mathcal{X}^n \subseteq \mathcal{X}^{n-1}$ we have that for all $n \geq m$ and $\mathbf{x} \in \mathcal{X}^n$ we have $x_i = T^m$. Once we have $S^n = \emptyset$ it means that all components of vectors in $\mathcal{X}^n$ are fixed hence $|\mathcal{X}^n| = 1$. According to lemma 3.2, this single vector in $\mathcal{X}^n$ is also max-min fair on $\mathcal{X}$.

q.e.d.

The algorithm presented in [49] for calculating the leximax minimal allocation is a particular

implementation of MP. The algorithm in each step maximizes the minimum rates of links, which is exactly step 4 of the MP algorithm, tailored to the problem considered. The overall complexity of the algorithm in [49] is thus the same as the complexity of MP. Since the feasible set considered there is compact convex, it follows from Theorem 3.1 and Proposition 3.3 that the leximax minimal allocation obtained in [49] is in fact a min-max fair allocation.

Numerical Examples

In order to illustrate the behaviour of MP, we consider two simple examples. The first one is the network from Figure 3.1. The set of feasible rates is

$$\mathcal{X} = \{(x_1, x_2) \,|\, 0 \leq x_1 \leq 7, 0 \leq x_2 \leq 3, x_1 + x_2 \leq 8\}, \tag{3.9}$$

and it is depicted on the right of Figure 3.1. We are looking for the max-min fair rate allocation.

In the first step of the algorithm we have $\mathcal{X}^0 = \mathcal{X}, \mathcal{R}^0 = \mathcal{X}, S^0 = \{1, 2\}$. By solving the linear program in step 4, we obtain $T^1 = 3$. We further have $\mathcal{X}^1 = \{(x_1, 3) \,|\, 3 < x_1 \leq 5\}, \mathcal{R}^1 = \{(x_1, 3) \,|\, 3 \leq x_1 \leq 5\}, S^0 = \{1\}$. Again by solving the linear program in step 4 we obtain $T^2 = 5$. Now we have $\mathcal{X}^2 = \emptyset, \mathcal{R}^2 = \{(5, 3)\}, S^2 = \emptyset$. The algorithm terminates and set $\mathcal{R}^2$ contains only the max-min fair rate allocation.

The second example we consider is the load distribution example from Figure 3.4. The set of feasible rates is

$$\mathcal{X} = \{(x_1, x_2) \,|\, 0 \leq x_1 \leq 7, 0 \leq x_2 \leq 3, 7 \leq x_1 + x_2 \leq 8\}, \tag{3.10}$$

and it is depicted on the right of Figure 3.4. We are looking for the min-max fair rate allocation.

In the first step of the algorithm we have $\mathcal{X}^0 = \mathcal{X}, \mathcal{R}^0 = \mathcal{X}, S^0 = \{1, 2\}$. By solving the linear program in step 4 (with minimization instead of maximization since we are looking for min-max fairness), we obtain $T^1 = 4$. We then have $\mathcal{X}^1 = \{(4, 3)\}, \mathcal{R}^1 = \{(4, 3)\}, S^0 = \emptyset$. The algorithm terminates and set $\mathcal{R}^2$ contains only the min-max fair rate allocation.

Note that when the max-min fair allocation does not exist, MP will not give one of the leximin maximal points, as one might expect. To see this, consider the examples from Figure 3.5. In both examples, in the first step of MP, we will have $T^1 = 1$ and $S^1 = \emptyset$, and the algorithm will return $(1, 1)$ as the optimal point. This point is neither leximin maximal, nor Pareto optimal. Therefore, one should be careful about using MP on problems that do not have max-min fair allocations.

3.4.2 The Water-Filling (WF) Algorithm

We now compare MP with the water-filling approach used in the traditional setting [8]. We here present a generalized version that includes minimal rate guarantees, as in [36].

We first need the concept of free disposal. It is defined in economics as the right of each user to dispose of an arbitrary amount of owned commodities [54], or alternatively, to consume fewer resources than maximally allowed. We then modify it slightly, as follows. Call $\mathbf{e}_i$ the unit vector $(\mathbf{e}_i)_j = \delta_{ij}$.

Definition 3.6 *We say that a set $\mathcal{X}$ has the free-disposal property if (1) there exists $\mathbf{m} \in \mathcal{X}$ with $x_i \geq m_i$ for all $\mathbf{x} \in \mathcal{X}$ and (2) for all $i \in \{1, \cdots, N\}$ and for all $\alpha \geq 0$ such that $\mathbf{x} - \alpha \mathbf{e}_i \geq \mathbf{m}$, we have $\mathbf{x} - \alpha \mathbf{e}_i \in \mathcal{X}$.*

Informally, free disposal applies to sets where each coordinate is independently lower-bounded, and requires that we can always decrease a feasible vector, as long as we remain above the lower bounds. We now describe the Water-Filling algorithm.

0. Assume $\mathcal{X}$ is free-disposal
1. let $S^0 = \{1, \cdots, N\}, \mathcal{X}^0 = \mathcal{X}, \mathcal{R}^0 = \mathcal{R}$ and $n = 0$
2. **do**
3. $n = n + 1$
4. Problem WF^n: maximize T^n subject to:
$$\begin{cases} (\forall i \in S^{n-1}) \quad x_i = \max(T^n, m_i) \\ \quad \text{for some} \quad \mathbf{x} \in \mathcal{X}^{n-1} \end{cases} \tag{3.11}$$
5. let $\mathcal{X}^n = \{\mathbf{x} \in \mathcal{X}^{n-1} \mid (\forall i \in S^{n-1})\, x_i \geq T^n, (\exists i \in S^{n-1})\, x_i > T^n\}$,
$\mathcal{R}^n = \{\mathbf{x} \in \mathcal{X}^{n-1} \mid (\forall i \in S^{n-1})\, x_i \geq T^n\}$
and $S^n = \{i \in \{1, \cdots, N\} \mid (\forall \mathbf{x} \in \mathcal{X}^n)\, x_i > T^n\}$
6. **until** $S^n = \emptyset$
7. return the only element in $\mathcal{X}^n$

When WF is equivalent to MP

The following theorem demonstrates the equivalence of MP and WF on free-disposal sets.

Theorem 3.4 *Let $\mathcal{X}$ be a max-min achievable set that satisfies the free-disposal property. Then, at every step n, the solutions to problems WF^n and MP^n are the same.*

Proof: Let us call T^1_{MP} the solution to the MP^1 and T^1_{WF} the solution to the WF^1. T^1_{WF} is obviously achievable in MP^1 so we have $T^1_{MP} \geq T^1_{WF}$. Suppose that $T^1_{MP} > T^1_{WF}$. This implies

that for all $s \in \{1, \cdots, N\}$ we have $(\mathbf{x}^1_{MP})_s \geq T^1_{MP}$. Due to the free-disposal property, we can successively decrease each of the components of $\mathbf{x}$ larger than corresponding lower bound in $\mathbf{m}$, until arriving to a vector $\mathbf{y}$, $y_i = \max(T^1_{MP}, m_i)$. This vector is feasible, which contradicts the optimality of T^1_{WF}. The same reasoning can be applied to the successive algorithm steps, by decreasing the dimension of the feasible set.

q.e.d.

Thus, under the conditions of the theorem, WF terminates and returns the same result as MP, namely the max-min fair vector if it exists. The theorem is actually stronger, since the two algorithms provide the same result at every step. However, if the free-disposal property does not hold, then WF may not compute the max-min fair allocation: see below Section 3.4.2 for such an example.

The examples previously mentioned of single path unicast routing [8], multicast util-max-min fairness [31, 78] and minimal rate guarantee [36, 40] all have the free-disposal property. Thus, the water-filling algorithm can be used, as is done in all the mentioned references. In contrast, the load distribution example [49] is not free-disposal, and all we can do is use MP, as is done in [49] in a specific example.

The multi-path routing example also has the free-disposal property, but the feasible set is defined implicitly. We discuss the implications of this in the next section.

Numerical Examples

In order to illustrate the behaviour of WF, we consider again the same two examples as in Section 3.4.1. In the first example, depicted in Figure 3.4, the feasible rate set, described by Equation (3.9), has the free-disposal property. It is easy to verify that sets $\{\mathcal{X}^i\}_{i=1\cdots3}, \{\mathcal{R}^i\}_{i=1\cdots3}, \{S^i\}_{i=1\cdots3}$ are taking exactly the same values as in the case of MP, described in Section 3.4.1. This confirms the findings of Theorem 3.4.

The second example we consider is the load distribution example depicted in Figure 3.4 and described by Equation (3.9). For this type of problem we cannot apriori set the upper limits in $\vec{m}$, as [40, 36], as they are not universal (they would need to depend on given network topology and are not known in advance). It is then easy to verify that the linear program in step 4 (with minimization instead of maximization since we are looking for min-max fairness) has no solution. Therefore, WF in this case cannot find the min-max fair rate allocation.

Note that the free-disposal property is a sufficient but not a necessary condition for MP to degenerate to WF. To see that, consider again the example from Figure 3.4. Suppose now $c_1 =$

$3, c_2 = 3, c_3 = 4$, and suppose in addition the minimum rate constraint is $x_1 + x_2 \geq 3$. The feasible rate set in this example has the same shape as in Figure 3.4, it is just translated to the left to touch the x_2 axis. In this case, it is easy to verify that the set still does not have the free-disposal property but WF does find the min-max allocation in a single step.

3.4.3 Complexity Of The Algorithms In Case Of Linear Constraints

Let us assume first that $\mathcal{X}$ is an n-dimensional feasible set defined by m linear inequalities. Each of the n steps of MP algorithm is a linear programming problem, hence the overall complexity is $O(nLP(n, m))$, where $LP(n, m)$ is the complexity of linear programming. The WF algorithm also has n steps, each of complexity $O(m)$ (since in step 4 we have to find the maximum value of T that satisfies the equality in each of the m inequalities, and take as the result the smallest of those) hence the complexity of WF is $O(nm)$. Linear programming has solutions of exponential complexity in the worst case, however in most practical cases there are solutions with polynomial complexity.

Assume next that $\mathcal{X}$ is defined implicitly, with an l-dimensional slack variable (for an example scenario, see multi path case on Figure 3.3). We can use MP, which works on implicit sets, and with complexity $O(nLP(n, m))$. If the set is free-disposal, we can also use WF, but we need to find an explicit characterization of the feasible set. In most of the existing problems, finding an explicit characterization of the feasible set can be done in polynomial time. To see that, consider again the example from Figure 3.3. The slack variables represent rates of different paths, and we are interested only in the end-to-end rates. Finding a set of feasible end-to-end rates is equivalent to a well known problem of finding maximum flows in a network [92] (see [49] for an example in the networking context). As shown in [92], this is a problem of a polynomial complexity. Note that it might be possible to construct an implicitly defined feasible set that cannot be converted to an explicit form in a polynomial time. However, we are not aware of any existing example of such a set, and a further analysis is out of the scope of our paper.

Once we have an explicit characterization, the remaining complexity of WF is still $O(mn)$. In practice, we are likely to be interested in finding the values of slack variables at the max-min fair vector; this itself is a linear program. However, it is sufficient to make the set explicit only once for a given problem. We conclude that in many practical problems, it is likely to be faster to make the set of constraints explicit and use WF rather than MP.

3.5 Example Scenarios

In this section we provide examples that arise in a networking context, which were not previously studied, and to which our theory applies. The first example deals with maximizing the life-time of one type of sensor network. The feasible set does have the free-disposal property, hence both WF and MP work. However, the goal of this example is to elucidate how to convert an implicit feasible set to an explicit one, which is illustrated here with an example.

The second and the third examples are taken from realistic problems that occur in P2P and wireless sensor networks, respectively. We show that in these two scenarios the feasible sets do not have the free-disposal property. We illustrate on simple but detailed numerical examples that WF does not work, whereas MP gives a correct result.

3.5.1 Minimum Energy Wireless Network

Our first novel example is the lifetime of a sensor network. Assume that we have a wireless sensor network, where an arbitrary set of source nodes collect and communicate information to corresponding destination nodes by relaying over intermediate nodes. Assume that each node has an initial energy, and it decreases each time it sends something. Once a node has no energy, it cannot send anything any more. Rather than analyzing the dynamics of the traffic, we consider the total amount of information that can be conveyed over the network during its lifetime.

Consider a network with $n = \{1 \cdots N\}$ nodes, and $l = \{1 \cdots L\}$ links. The initial energy of node n is E_n. The energy needed to send one bit over link l is e_l. There is a set of $s = \{1 \cdots S\}$ source destination pairs that communicate data. We denote with X_s the total amount of bits transferred over the network during its lifetime by source-destination pair s. Those bits may be transferred over one of $p = \{1 \cdots P\}$ different paths, and the number of bits transferred over path p is denoted with Y_p. We further say $a_{n,l}$ is 1 if a link l contains node n, else is 0, and $b_{l,p}$ is one if path p contains link l.

In a general case, a feasible information allocation can be described with the following constraint set

$$\mathcal{X} = \{X_s \,:\, E_n \geq \sum_l a_{n,l} e_l \sum_p b_{l,p} Y_p,\; X_s = \sum_p d_{s,p} Y_p\},$$

Set $\mathcal{X}$ is convex, thus it is also max-min achievable. It also has the free-disposal property so max-min fair allocation can be found by either WF or MP.

A simple four node example is given in Figure 3.8. Source S_1 sends data to destination D_1, and source S_2 sends data to destination D_2 (which happens to be the same node as D_1). Source

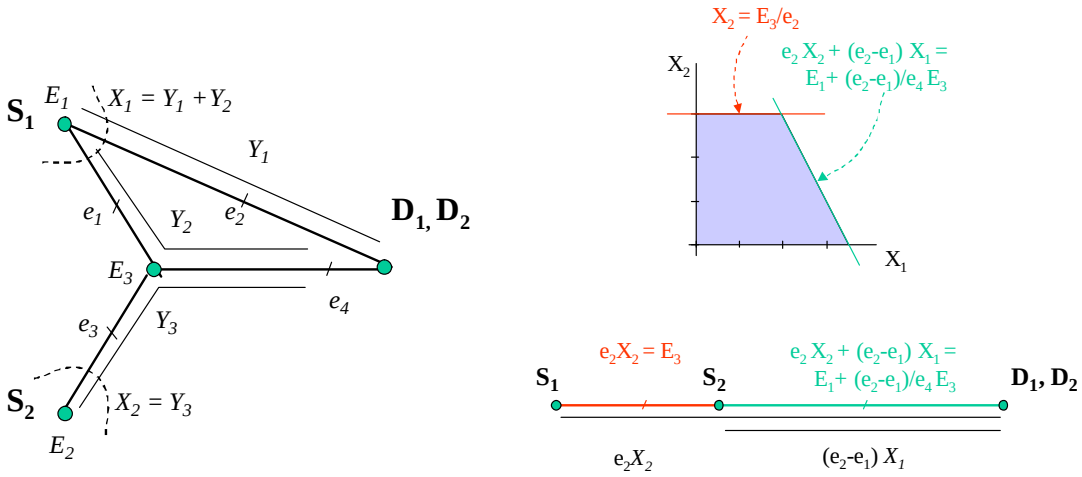


Figure 3.8: Minimum energy networks: Left: a four node example, with two sources and two destinations. Upper right: feasible set for the example on the left. Lower right: Single path routing problem with the equivalent feasible set.

S_1 sends over two paths, a direct path Y_1, and path Y_2 over node 3. Source S_2 sends over one path Y_3, over node 3. Initial energy of a node n is E_n, and a cost of sending each bit over link l is e_l, hence we have a following set of constraints

$$E_1 \geq e_2 Y_1 + e_1 Y_2, \tag{3.12}$$

$$E_2 \geq e_3 Y_3, \tag{3.13}$$

$$E_3 \geq e_4 (Y_2 + Y_3), \tag{3.14}$$

$$X_1 = Y_1 + Y_2, \tag{3.15}$$

$$X_2 = Y_3, \tag{3.16}$$

where X_i is the total amount of information transferred from S_i to D_i. We are looking for a max-min fair allocation of information transferred.

In order to decrease complexity of WF, we have to find an explicit characterization of the feasible set, hence we need to get rid of the slack variables and to define constraints solely on X_1, X_2. To do this, one can find max flows, using e.g. augmenting path method [92], for flows from S_1 to D_1, from S_1 to D_1, and jointly from S_1 and S_2 to D_1 and D_2. In the specific example from Figure 3.8, assuming in our simple example that $E_1 > E_3$ and $e_1 > e_2$, we can transform feasible set (3.12) - (3.16) into

$$X_2 \leq E_3/e_2, \tag{3.17}$$

$$e_2 X_1 + (e_1 - e_2) X_2 \quad \leq \quad E_1 + E_3(e_2 - e_1)/e_4. \tag{3.18}$$

as depicted in Figure 3.8, upper right, and we can apply WF to find max-min fair allocation. The same approach can be extended straightforward to an arbitrary network topology and performed in polynomial time; see Section 3.4.3 for discussion. It is interesting to notice that there exists a single-path routing problem with the same feasible set, shown in Figure 3.8, lower right, where capacities of links correspond to inequalities Equation (3.17) and Equation (3.18). Therefore we can still identify bottlenecks in the classical sense, although now they do not have a clear physical implementation (they correspond to maximum cuts in the max-flow min-cut algorithm).

This model is an extension of the model presented in [14]. In [14] they maximize only the shortest lifetime, hence solving only MP^1. This is not useful in the case of large networks, where a death of one node cannot be treated as a death of the network itself. Also, in [14] they optimize the lifetime given the traffic matrix, whereas we allow an arbitrary traffic matrix and we optimize the amount of information transferred. Thus, [14] is a special case of our model presented here.

On the contrary, our analysis is focused only on properties of the performance metric, and not on a specific realization of an algorithm. Unlike [14], where a fully distributed algorithm is given, we only give a centralized solution, and an implementation of a distributed protocol is out of scope and remains as a future work.

3.5.2 Load Distribution In P2P Systems

Let us consider a peer-to-peer network, where several servers can supply a single user with parts of a single data stream (e.g., by using Tornado codes [39]). There is a minimal rate a user needs to achieve, and there is an upper bound on each flow given by a network topology and link capacities.

Let $\mathbf{x}$ be the total loads on servers, $\mathbf{y}$ the flows from servers to clients, $\mathbf{z}$ the total traffic received by clients, $\mathbf{c}$ the capacities of links and $\mathbf{m}$ the minimum required rates of the flows. We can then represent the feasible rate set as

$$\mathcal{X} = \{\mathbf{x} : (\exists \mathbf{y}, \mathbf{z})\ A\mathbf{y} \leq \mathbf{c}, B\mathbf{y} = \mathbf{x}, C\mathbf{y} = \mathbf{z}, \mathbf{z} \geq \mathbf{m}\}, \quad A, B, C \geq 0, \tag{3.19}$$

where A, B and C are arbitrary matrices defined by network topology and routing.

A simple example depicted in Figure 3.4. Client D receives data from both servers S_1 and S_2 and it wants minimal guaranteed rate m. There is flow y_1 going from S_1 to D over links 1 and 3, and flow y_2 going from S_2 to D over links 2 and 3. We have that the total egress traffic of S_1 is

$x_1 = y_1$, and from S_2 is $x_2 = y_2$. The total ingress traffic of D is $z_1 = y_1 + y_2$. We thus have the following matrices

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}, B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, C = \begin{bmatrix} 1 & 1 \end{bmatrix},$$

that define the constraint set, visualized in Figure 3.4.

In a peer-to-peer scenario, each server is interested in minimizing its own load, hence it is natural to look for the min-max fair vector on set $\mathcal{X}$, which minimizes loads on highly loaded servers.

Set $\mathcal{X}$ is convex, hence it is the min-max achievable. It does not have the free-disposal property in general, and WF is not applicable. This is showed in Section 3.4.1 on a simple example. Min-max fair allocation can be found by means of the MP algorithm. This is illustrated on the example in Section 3.4.1.

Note that this form of a feasible set is unique in that it introduces both upper and lower bounds on a sum of components of $\mathbf{x}$ and, as such, is more general than the feasible sets in the above presented examples, such as [49].

3.5.3 Maximum Lifetime Sensor Networks

In this section we consider an example motivated by [14, 77]. This example is somewhat similar to the example in Section 3.5.1. However, Section 3.5.1 considers wireless networks where the underlying physical interface eliminates intra-node interference. This is the case only for a small class of physical and MAC layers (i.e. a model of 802.11 where the interference is eliminated by exclusions). In a more realistic example of a physical model, a transmission on one link provokes interference, which decreases the rates of other links in a region. In this example we assume that the network is built on the top of the ultra-wide band physical model described in [96].

Consider a set of $n = \{1 \cdots N\}$ nodes, some of which are sensors and some are sinks. We assume sensors feed data to sinks over the network, and can do this by sending directly, or relaying over other sensors or sinks. When node s sends data to node d, it does so using some transmission power P_s. The signal attenuates while propagating through space, and is received at d with power $P_s h_{sd}$, where h_{sd} is called the attenuation between s and d, and is an arbitrary positive number.

Receiver d tries to decode the information sent by s in presence of noise and interference. If N denotes the white background noise, than the total interference experienced by D is $I =$

$N + \sum_{i \neq s} P_i h_{id}$. The maximum rate of information d can achieve is then [96]

$$x_{sd} = K \frac{P_s h_{sd}}{N + \sum_{i \neq s} P_i h_{id}}.$$

We also assume that a node can only send to or receive from one node at a time.

In addition, nodes can change transmission power in time. We assume a slotted protocol, where in every slot t, every node s can choose an arbitrary transmission power $P_s(t)$ (if s chooses not to transmit, it sets $P_s(t) = 0$). Link sd will then achieve rate $x_{sd}(t)$ that depends on allocated powers, as explained above. We denote with x_{sd}^{AVG} the average rate of link sd throughout a schedule. Let $\mathbf{x}^{AVG}$ be the vector of all $\{x_{sd}^{AVG}\}_{1 \leq s,d \leq N}$. We denote by $\mathcal{X}$ a set of feasible $\mathbf{x}^{AVG}$, that is such that there exists a schedule and power allocations that achieve those rates. Similarly to the average rate, we can calculate the average power a node dissipates during a schedule, which we denote with P_s^{AVG}. We denote with $\mathcal{P}(\mathbf{x}^{AVG})$ a set of possible average power dissipations one can use to achieve average rate $\mathbf{x}^{AVG}$. For a more detailed explanation of the model, see Chapter 5 and Chapter 6.

From the application point of view, we assume sensors measure the same type of information. Each of the several sinks needs to receive a certain rate of the information, regardless from what sensor it comes. Let us denote with R_d the total rate of information received by sink d. We then have a constraint $R_d \leq M_d$.

In order to define routing, let us further introduce a concept of paths. Path $p = \{1 \cdots P\}$ is a set of links. We say $A_{l,p} = 1$ if link $l = (s, d)$, for some s, d, belongs to path p. Otherwise, $a_{p,l} = 0$. We also say $B_{s,p} = 1$ and $C_{s,p} = 1$ if node s is the starting or the finishing point of the path p, respectively. Let y_p be the average rate on path p.

The goal is to minimize average power dissipations, under the above constraints. The set of feasible average power dissipations can be formally described as

$$\mathcal{P} = \{\mathbf{p}^{AVG} \,|\, (\exists \mathbf{x}^{AVG} \in \mathcal{X})\, \mathbf{p}^{AVG} \in \mathcal{P}(\mathbf{x}^{AVG}), A\mathbf{y} \leq \mathbf{x}^{AVG}, \mathbf{R} = C\mathbf{y} \leq \mathbf{M}\}$$

We are interested in finding the min-max average power allocation over set $\mathcal{P}$.

As one might expect, this is an extremely difficult problem to solve for an arbitrary network and has not been fully solved, to the best of our knowledge. However, we want to illustrate in a simple example from Figure 3.9, that the feasible set does not always have the free-disposal property, and furthermore that WF, as such, cannot be used.

In our simple example from Figure 3.9, we consider two sensors, S_1 and S_2, and two sinks, D_1 and D_2. We have three links, (S_1, D_1), (S_2, D_1), (S_2, D_2), and three paths that coincide with

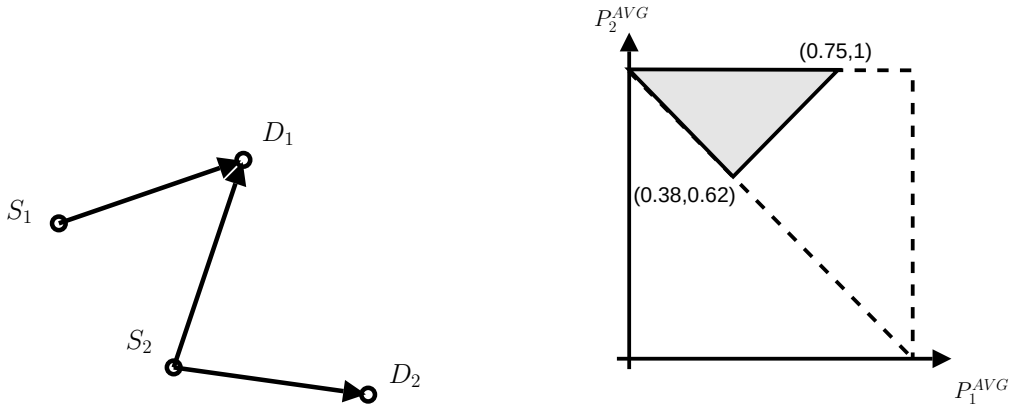


Figure 3.9: Sensor example: On the left an example of a network with 2 sensors and 2 sinks is given. We let $P^{MAX} = N = 1$, and $h_{S_1D_1} = h_{S_1D_2} = 1, h_{S_2D_1} = 10, h_{S_2D_2} = 0.7$, and the lower bounds on rates are $M1 = 0.6, M2 = 0.4$. On the right, the set of feasible average power dissipations is given.

each link (we assume other links cannot be established due to for example a presence of a wall).

It is shown in Chapter 6 that in this type of network any average rate allocation can be achieved by using the following simple power allocation policy: when a node is transmitting, it does so with maximum power; otherwise it is silent. It follows that any possible schedule in the network can have four possible slots:

Slot 1 of duration α_1: Only sensor S_1 sends to sink D_1 with full power P^{MAX}. S_2 is silent.

Slot 2 of duration α_2: Sensor S_1 sends to D_1 while S_2 sends to D_2.

Slot 3 of duration α_3: Only S_2 sends to D_1.

Slot 4 of duration α_4: Only S_2 sends to D_2.

If we normalize the duration of the schedule, we have $\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = 1$.

Under the above scheduling, we have the following average rates and average dissipated powers

$$R_1 = \alpha_1 \frac{P^{MAX} h_{S_1D_1}}{N} + \alpha_2 \frac{P^{MAX} h_{S_1D_1}}{N + P^{MAX} h_{S_2D_1}} + \alpha_3 \frac{P^{MAX} h_{S_2D_1}}{N}, \tag{3.20}$$

$$R_2 = \alpha_2 \frac{P^{MAX} h_{S_2D_2}}{N + P^{MAX} h_{S_1D_2}} + \alpha_4 \frac{P^{MAX} h_{S_2D_2}}{N}, \tag{3.21}$$

$$P_1^{AVG} = (\alpha_1 + \alpha_2 + \alpha_3) P^{MAX}, \tag{3.22}$$

$$P_2^{AVG} = (\alpha_2 + \alpha_4) P^{MAX}. \tag{3.23}$$

The set of feasible average powers is thus $\mathcal{X} = \{(P_1^{AVG}, P_2^{AVG}) \,|\, (\exists \alpha_{1\cdots 4}) \sum_{i=1}^{4} \alpha_i = 1, R_1 \geq M_1, R_2 \geq M_2\}$.

In order to obtain a numeric example, we take $K = P^{MAX} = N = 1, h_{S_1D_1} = h_{S_1D_2} = 1, h_{S_2D_1} = 10, h_{S_2D_2} = 0.7$, and $M1 = 0.6, M2 = 0.4$. Plugging these values in (3.20)-(3.23) and simplifying the constraints, we achieve the following set of inequalities that defines set $\mathcal{X}$:

$$\begin{aligned} P_1^{AVG} + P_2^{AVG} &\geq 1, \\ P_1^{AVG} + \alpha_3 &\leq 1, \\ 7P_1^{AVG} + 14\alpha_3 + 1 &\leq 7P_2^{AVG}, \\ P_1^{AVG} + 110\alpha_3 - 3.4 &\geq 10P_2^{AVG}, \\ P_1^{AVG}, P_2^{AVG}, \alpha_3 \in [0,1]. & \end{aligned}$$

The set $\mathcal{P}$ is depicted on the right of Figure 3.9. It is easy to verify that this set does not have the free-disposal property. We verify that the first step of WF algorithm has no solution, hence water filling does not give the min-max allocation. On the other hand, a single iteration of MP gives us the min-max allocation on the set $\mathcal{X}$ which in this case is $(0.38, 0.62)$. We underline again that only due to the simplicity of the example, WF fails at the first step, and MP solves the problem in one step. In a more complex example WF might fail on an arbitrary step whereas MP will again solve the problem in more than one step. However, due to the simplicity of the presentation we give here only a 4 node example.

3.6 Summary of Main Findings and Conclusions

The main findings in this chapter are:

- Our first finding is on the existence of max-min fairness. We gave a large class of continuous sets on which a max-min fair allocation does exist, and we theoretically proved the existence. This class contains the compact, convex sets of $\mathbb{R}^N$, but not only. We also illustrated in a few examples that there are sets in which max-min fairness does not exist, and thus that our result is not trivial.

- Our second finding is on algorithms that locate the max-min fair allocation. In Section 3.4, we gave the general purpose, centralized algorithm called *max-min programming (MP)*, and we proved that it finds the max-min fair allocation in all cases where it exists. Its

complexity is of the order of N linear programming steps in $\mathbb{R}^N$, in general, whenever the feasible set is defined by linear constraints.

- The third finding is on the relation between the general MP algorithm and the existing WF algorithm. We recalled the definition of the free-disposal property and showed that, whenever it holds, max-min programming (MP) degenerates to the simpler water-filling (WF) algorithm (originally defined in [8]), whose complexity is much less. The free-disposal property corresponds to the cases where a bottleneck argument can be made, and water-filling is the general form of all previously known centralized algorithms for such cases. We noted that WF requires the feasible set to be given in explicit form, unlike MP, and we discussed the case of an implicit feasible set with the free-disposal property.

We presented a novel approach to analyzing properties of max-min fairness. Instead of considering a specific networking problem with an underlying network topology, we focused only on the feasible rate sets. Therefore, our framework does not depend on a specific problem; it is general and it unifies the existing approaches that analyze max-min fairness.

We showed the application of our findings in three networking examples: One example is on minimum-energy wireless networks; The second is on load distribution in P2P systems; The last example is maximum-lifetime sensor networks. We gave specific, numerical examples where max-min fair allocation exists but the feasible sets do not have the free-disposal property, hence a classical water-filling cannot be used. We showed in these examples how MP finds max-min fair allocation even when the free-disposal does not hold. This way, we verify that our framework unifies previous results and extends the applicability of max-min fairness to new scenarios.

All our results are given for max-min fairness; they apply mutatis mutandis to min-max fairness. They are valid for weighted and unweighted max-min and min-max fairness, using the transformation given in Section 3.3.1. Distributed algorithms for the computation of max-min fair allocations [23, 15] are left outside the scope of this analysis.

Chapter 4

Rate Performance Objectives of Multi-hop Wireless Networks

4.1 Introduction

In this chapter we consider the question of which performance metric to maximize when designing ad-hoc wireless network protocols such as routing or MAC. We focus on maximizing rates under battery-lifetime and power constraints.

Commonly used metrics are total capacity (in the case of cellular networks) and transport capacity (in the case of ad-hoc networks). However, it is known in traditional wired networking that maximizing total capacity conflicts with fairness, and this is why fairness oriented rate allocations, such as max-min fairness, are often used.

We review this issue for wireless ad-hoc networks. Indeed, the mathematical model for wireless networks has a specificity that makes some of the findings different than in the wired case. It has been reported in the literature on Ultra Wide Band that gross unfairness occurs when maximizing total capacity or transport capacity . We confirm by a theoretical analysis that this is a fundamental shortcoming of these metrics in wireless ad-hoc networks, as it is for wired networks.

The story is different for max-min fairness. Although it is perfectly viable for a wired network, it is much less so in our setting. We show that, in the limit of long battery-lifetimes, the max-min allocation of rates always leads to strictly equal rates, regardless of the MAC layer, network topology, channel variations, or choice of routes and power constraints. This is due to the "solidarity" property of the set of feasible rates. The consequence of the phenomenon is that all flows receive the rate of the worst flow, and this leads to severe inefficiency. We show

numerically that the problem persists even when battery-lifetime constraints are finite. This generalizes the observation reported in the literature that, in heterogeneous settings, 802.11 allocates the worst rate to all stations, and it shows that this is inherent to any protocol that implements max-min fairness.

Utility fairness is an alternative to max-min fairness, which approximates rate allocation performed by TCP in the Internet. We analyze, by numerical simulations, different utility functions and we show that the proportional fairness of rates or transport rates, a particular instance of utility based metrics, is robust and achieves a good trade-off between efficiency and fairness, unlike total rate or maximum fairness. We thus recommend that metrics for the rate performance of mobile ad-hoc networking protocols be based on proportional fairness.

We define a model of a wireless multi-hop network, and we use it as a tool to analyze efficiency and fairness of different design criteria for various network technologies, and to prove the above findings. Our model allows for the most general assumptions on a physical layer (including variable rate 802.11, UWB or CDMA), MAC and routing protocols. And for a given network topology, channel and noise statistics and traffic demand, we characterize a set of feasible long-term average end-to-end rate and transport rate allocations. Next, we find the optimal allocations on the two sets, with respect to the three design criteria considered. In some numerical examples, where it is not possible to find an exact solution of the optimization problem due to its non-convexity, we consider an approximation that is close to the optimal solution and that allows us to accurately characterize the efficiency and fairness of the optimum.

The outline of the chapter is as follows. In the next section we give descriptions of different rate-based performance metrics and their properties. In Section 4.3, we describe system assumptions. In Section 4.4 we give a mathematical formulation of the model of a network. In Section 4.5, Section 4.6 and Section 4.7 we present findings related to max-min fairness, maximizing total capacity and proportional fairness objectives, respectively. In Section 4.8 we discuss the influence of long-term average power constraints. In the last section we summarize the main findings and give conclusions.

4.2 Rate-based Performance Metrics with Power Constraints

We consider the question of what metric to use when evaluating the performance of ad-hoc wireless network protocols such as routing or MAC. We focus on maximizing rates under battery-lifetime and power constraints. Typical application examples are networks of wireless laptops and PDAs; this is also the framework used by many papers analyzing various models of physical

layers (purely information theoretic approach [85, 90], CDMA [19], UWB [21]). In contrast, some sensor networks put more emphasis on minimizing energy under minimum rate constraints. Here, we study the former and leave the latter as a future work.

For cellular wireless networks, a frequently chosen performance metric is total capacity, i.e. the sum of the rates of all flows. An extension that maximizes a weighted sum of the rates is applied in CDMA/HDR [67]. In multi-hop wireless networks, the same metrics are used, but also used is transport capacity, a variant popularized by Gupta and Kumar in [30]. This is in fact a weighted sum of rate, where weights are the distances between the source and the destination of each flow.

4.2.1 The Tension Between Efficiency and Fairness

Traditionally, wired networking has also focused on performance metrics that incorporate some form of fairness. Indeed, it is known that considering only total capacity yields gross unfairness if implemented in a wired network [51]. The unfairness has been observed in the framework of multi-hop ad-hoc networks in [21]. Hence, different performance metrics that account for fairness have been developed. A typical example is max-min fairness [8], analyzed in detail in Section 3, which is used in many existing networking protocols, including the ABR mode of ATM [15]. This is an egalitarian approach by which the rate of a flow can be increased only when it is not possible to increase the rate of an already smaller flow. Max-min fairness is often viewed as an extreme fairness (as discussed in Section 3.3.4); this justifies using a *fairness index* that measures the departure from max-min fairness (it is a slight variant of the fairness index defined by Jain in [41]; see Section 4.4.4).

Max-min fairness is also used, often implicitly, in many existing wireless multi-hop network protocols (e.g. [86, 37]). In fact, as we show, 802.11 essentially implements max-min fairness. However, in wireless networks, there is still no tradition of evaluating a system in light of both total rate and fairness. It turns out that the issue of the trade-off between efficiency and fairness is significantly different in wireless than in wired networking, due to the peculiarities of the mathematical models for wireless networks. This phenomenon has been first observed in [7]. In particular, we find that the allocations that implement max-min fairness have fundamental efficiency problems. This is due to the "solidarity" property of the set of feasible rates (Section 4.5.1).

Another way to reduce the tension between efficiency and fairness is to use a weighted sum of the rates as a design objective. The most well-known example of this type of criteria in wireless networks is transport capacity [30] where each flow is assigned a weight equal to the distance

between the source and the destination of the flow. We show in Figure 4.7 in Section 4.7 that this approach does not reconcile the tension.

4.2.2 Utility Fairness

Utility fairness is often used as an alternative, a less egalitarian approach to max-min fairness. It corresponds to the "utility" metric $\sum_j U(x_j)$ where x_j is the rate of flow j and $U()$ is a concave function (called the utility function). The concept of utility is a convenient way to represent user preferences and a utility function $U()$ can be interpreted as a user satisfaction [54]. The Internet congestion control performed by TCP approximates some form of utility fairness.

The properties of utility fairness depend on the choice of utility functions. It is known in wired networking that, for a large class of strictly concave utility functions, maximizing the utility metric is fairer than maximizing the total capacity , but less egalitarian than a max-min fair allocation.

The most often used class of utility functions is of form

$$U(x,\xi) = \begin{cases} (1-\xi)^{-1}x^{1-\xi} & \text{if } \xi \neq 1 \\ \log(x) & \text{if } \xi = 1 \end{cases}, \tag{4.1}$$

proposed in [59]. This class of utility function is general enough to incorporate the most often used objectives: rates maximization (for $\xi = 0$), *proportional fairness* [47] (for $\xi = 1$), *minimum potential delay* [55] (for $\xi = 2$) and max-min fairness (for $\xi \to \infty$). We consider this form of utility in order to evaluate if and how the trade-off between efficiency and fairness can be tuned through parameter ξ of the utility function.

Note that the utility approach can easily be extended to account for power and energy – not in the form of constraints as we do here, but through a cost function subtracted from the utility metric. This is explored for example by Baldi et al. [4]. We leave such metrics out of the scope of this chapter, as we focus on rate-based metrics with power constraints.

4.3 System Assumptions and Modelling

We analyze an arbitrary ad-hoc wireless network that consists of a set of nodes, and each two nodes that directly exchange information are called a link. Next we give properties of the physical model of communications on links.

4.3.1 Physical Model Properties

Power, Attenuation and White Noise

For each pair of nodes, we define a signal attenuation (or fading). If a signal on link l is transmitted with power P_l, then it is received with power $P_l h_l$, where h_l is the attenuation on link l. This attenuation is usually a decreasing function of a link length due to power spreading in all directions.

The attenuation is also random, due to mobility and variability in the transmission medium. We assume the system is in a state that varies over time according to some stationary stochastic process $\{(S(t)), t > 0\}$, where $S(t) \in \mathcal{S}$ is the state at time t. The attenuation of link l is a function $h_l(S(t))$ of the state of the system.

We assume the network is located on a finite surface and that all attenuations are always strictly positive ($h_l(s) > 0$ for all $s \in \mathcal{S}$), hence every node can be heard by any other node in the network at all times, and there is no clustering.

We also assume the presence of a white Gaussian background noise in the system. This noise may differ in space and time. Similar to attenuations, we model it as a function of the system state: $\eta_l(s)$ is the white noise on link l when system state is s ($\eta_l(s) > 0$ for all $s \in \mathcal{S}$).

Rate Function

All physical links are point-to-point, that is, each link has a single source and a single destination. There are more advanced models such as broadcast channel and relay channel [18] that attain higher performances, but they are not used in most of the contemporary networks, and their performances are in general not known and are still an open research issue. Broadcast is used by 802.11, by MAC layer control packets, but this is an aspect we do not model here.

A node can either, at one given time, send to one destination or receive from one source. There are more complex transmitter or receiver designs that can overcome these limitations. An example is a multi-user receiver that could receive several signals at a time. This would change the performance of links with a common destination, but would not change the interactions over a network. However, these more complex techniques are not used in contemporary multi-hop wireless networks (like 802.11, UWB, bluetooth or CDMA) due to high transceiver complexity, and we do not consider them here.

We model rate as a *strictly increasing* function $r(\mathrm{SINR})$ of the signal-to-interference-and-noise ratio at the receiver, which is a ratio of received power by the total interference perceived by the receiver including the ambient noise and the communications of other links that occur at

the same time. This model corresponds to a large class of physical layer models, for example:

- Shannon capacity of a Gaussian channel [18]:
  $r(\mathrm{SINR}) = 1/2 \log_2(1 + \mathrm{SINR})$.
- Win-Scholts model of Ultra-wide band [21]: $r(\mathrm{SINR}) = K \times \mathrm{SINR}$.
- CDMA HDR [67]: $r(\mathrm{SINR})$ is a stair function of SINR.
- variable rate 802.11 [35]: $r(\mathrm{SINR})$ is a stair function of SINR.

We note that in the last two models, rate is not a strictly increasing function of SNR but in most applications can be approximated as such. In contrast, in the basic model of 802.11 (e.g. [30]), the rate is assumed to be constant hence this model does not fit this framework.

4.3.2 MAC Protocol

We further assume a slotted protocol. In each slot a node can either send data, receive, or stay idle, according to the rules defined in Section 4.3.1. Each slot has a power allocation vector associated with it, which denotes what power is used for transmitting by the source of each link. If a link is not active in a given slot, its transmitting power is 0. An example of a schedule is depicted on Figure 4.1.

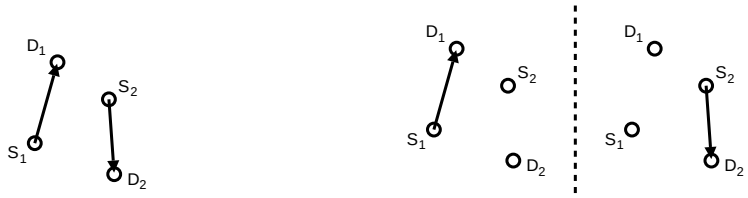


Figure 4.1: Two examples of schedules: consider a network with two links, (S_1, D_1) and (S_2, D_2). One possible schedule, depicted on the left, is to have both links sending all the time. Another schedule is to have two slots. In the first slot link (S_1, D_1) transmits and in the second slots link (S_2, D_2) transmits.

For each slot we define its relative frequency, which represents a fraction of time of the schedule it occupies. Consider the example from the right of Figure 4.1. There, we have two slots, slots 1 and 2. Suppose we first schedule slot 2 during a time interval T_1^2, then we schedule slot 1 during a time interval T_1^1, then again slot 2 for another T_2^2 and finally slot 1 for another T_2^1. This schedule then repeats in time. The relative frequency of slot 1 is thus $(T_1^1 + T_2^1)/(T_1^1 + T_2^1 + T_1^2 + T_2^2)$ and the relative frequency of slot 2 is $(T_1^2 + T_2^2)/(T_1^1 + T_2^1 + T_1^2 + T_2^2)$.

In general, although durations of slots may vary depending of the available data at nodes, or packet sizes, the average link rate depends only on the relative frequency. A schedule can consist of an arbitrary number of slots of arbitrary durations and we consider only frequencies of the slots and power allocations within slots as a parameter of a schedule.

There is a schedule assigned to every possible system state, and these parameters are a function of the system state. Examples of such scheduling protocols are opportunistic schedulers [90, 81]: depending on the actual link fadings, the protocol is very likely to schedule the link with the best channel conditions.

We assume an ideal MAC protocol that always knows the state of the system and disposes of an ideal control plane with zero delay and infinite throughput to negotiate power allocations. Given the state, it calculates the optimal transmission power of each link in each slot, as well as the optimal slot frequencies, in a centralized manner and according to a predefined metric. It then schedules a random slot from the schedule that corresponds to the system state, with the probability equal to the relative frequencies of the slots (as in [66]).

The assumptions on a MAC protocol, with an ideal control plane and the complete knowledge of the system, might seem overly simplified for a xswireless multi-hop network scenario. However, as shown in [66, 19], there exist decentralized protocols whose performances come close to the optimal one. Also, by considering an ideal protocol, we focus our analysis on the properties of performance metrics, and not artifacts of leaks in protocol design. Our assumption corresponds to neglecting the overhead (in rate and power) of the actual MAC protocol.

4.3.3 Routing Protocol and Traffic Flows

We assume an arbitrary routing protocol. Flows between sources and destinations are mapped to paths, according to some rules specific to the routing protocol. At one end of the spectrum, nodes do not relay and only one-hop direct paths are possible. At the other end, nodes are willing to relay data for others and multi-hop paths are possible. There can be several parallel paths. The choice of routes may also depend on the current state of the system. All these cases correspond to different constraint sets in our model, as explained in Section 4.4.1.

We assume all nodes always have data to send. Flow control is assumed to use all available links' capacities in order to maximize performance while keeping the network stable without dropping packets. Hence, it can be TCP, UDP (with with optimally choosen rates) or some other best-effort flow control.

4.3.4 Power Constraints

There are three types of power constraints in a wireless network: peak constraint, short-term average constraint and long-term average constraint. Here we describe them in detail:

Peak power constraint: Given a noise level on a receiver, a sender can decide which codebook it will use to send data over the link during one time slot. Different symbols in the codebook will have different powers. The maximum power of a symbol in a codebook is then called *the peak power*. It depends on the choice of the physical interface and its hardware implementation and we cannot control it. It limits the choice of possible codebooks, and it puts restrictions on the available rate. For example, the rate of an UWB link, given the average SNR on the receiver, depends on the shape of the pulse, thus on the peak power level of the pulse [96]. In our model, the peak power constraint is integrated in a rate function, given as an input.

Transmission power : We assume a slotted system. In each slot a node chooses a codebook and its average power, and sends data using this codebook within the duration of the slot. We call *the transmission power* the average power of a symbol in the codebook. This is a short-term average power within a slot, since a codebook is fixed during one slot. We assume that this transmission power is upper-bounded by P^{MAX}. This power limit is implied by technical characteristics of a sender and by regulations, and is not necessarily the same for all nodes. For example, this is the only power constraint that can be set by users on 802.11 equipment.

Long-term average power : While transmitting a burst of data (made of a large number of bits), a node uses several slots, and possibly several different codebooks. Each of these codebooks has its transmission power. We call *the long-term average power* the average of transmission powers during a burst, and we assume it is limited by $\overline{P}^{MAX}$. Long-term average power is related to the battery lifetime in the following way:

$$T_{\text{lifetime}} \approx \frac{E_{\text{battery}}}{\overline{P}^{MAX} \times u}$$

where T_{lifetime} is the battery lifetime, E_{battery} is the battery energy, $\overline{P}^{MAX}$ is the long-term average power constraint and u is the fraction of time a node has data to send (or activity factor, measured in Erlangs). The approximation corresponds to ignoring overhead spent managing the sleep / wakeup phases, etc. $\overline{P}^{MAX}$ is thus set by a node to control its lifetime; it can vary from node to node.

We incorporate explicitly in our model the transmission power and the long-term average power constraints. The peak power is incorporated implicitly through the choice of the rate function.

4.4 Mathematical Formulation of The Feasible Sets and of the Metrics

4.4.1 Feasible Set of Rates

We model the wireless network as a set of I flows, L links, O nodes and N time-slots.

At every time instant, the system is in some state belonging to the set of possible states $\mathcal{S}$ (depicting different channel conditions and noise levels). There is a finite (possibly very large) number of states $|\mathcal{S}|$. Every flow can use one or several paths. There are P paths ($P \leq 2^L$). We also assume that there exists a schedule for each system state, and that the schedule consists of time slots $n = 1, \cdots, N$ of relative frequency $\alpha_n(s)$. We are interested in the set of feasible long-term average end-to-end rates.

Here, we give a list of notations used throughout this, and the following chapters:

- $S \in \mathcal{S}$ is a random variable representing the system state. We assume system state is a continuous, stationary random process, and due to stationarity, we drop the time argument (the model can be easily extended to a larger class of ergodic processes, as in [66, 9]).
- $h_{l_1 l_2}(s)$ is the attenuation of a signal from the source of link l_1 to the destination of link l_2 when the network is in state s.
- $\alpha_n(s)$ for $n \in \{1, \cdots, N\}$ is the relative frequency of time slots in the schedule assigned to system state s. We have $\sum_{n=1}^{N} \alpha_n(s) = 1$.
- $\eta_l(s)$ is the white noise on link l when system state is s ($\eta_l(s) > 0$ for all $s \in \mathcal{S}$).
- $\mathbf{p}_l^n(s)$ is the power scheduled to link l in slot n and state s.
- $\mathbf{SINR}_l^n(s)$ is the signal-to-interference-and-noise ratio at the receiver of link l in slot n and system state s.
- $r(\mathrm{SINR})$ is the rate function that defines the rate of a link given the signal-to-interference-and-noise ratio at the receiver of the link.
- $\mathbf{x}_l^n(s) = r(\mathbf{SINR}_l^n(s))$ is the rate of link l in slot n and state s.
- $\bar{\mathbf{x}}_l(s) = \sum_{n=1 \cdots N} \alpha_n(s) \mathbf{x}_l^n(s)$ is the average rate achieved on link l in state s.
- $\mathbf{y}(s) \in \mathbb{R}^P$ is the vector of average rates used on paths in state s.

- $\mathbf{f}(s) \in \mathbb{R}^I$ is the vector of average rates achieved by flows when the network is in state s,
- $F(s)$ (flow matrix) is such that $F_{i,p}(s) = 1$ if path p belongs to flow i in the state s, and 0 elsewhere. We have $\mathbf{f}(s) = F(s)\mathbf{y}(s)$.
- $R(s)$ (routing matrix) is such that $R_{l,p}(s) = 1$ if path p uses link l in state s. We have $\bar{\mathbf{x}}(s) \geq R(s)\mathbf{y}(s)$. The matrix $R(s)$ is defined by the routing algorithm and may depend on the state s.
- $\bar{\mathbf{f}} \in \mathbb{R}^I$ is the vector of long-term average rates achieved by flows. We have $\bar{\mathbf{f}} = \mathbb{E}\{\mathbf{f}(S)\}$.
- $\bar{\mathbf{p}}_l = \mathbb{E}\{\sum_{n=1\cdots N} \alpha_n(s)\mathbf{p}_l^n(s)\}$ is the long-term average power consumed by link l.
- $\mathbf{P}_l^{MAX} \geq \mathbf{p}_l^n(s)$ (for all n, s), are the transmission power constraints.
- $\overline{\mathbf{P}}_l^{MAX} \geq \bar{\mathbf{p}}_l$, are the long-term average transmission power constraints.
- $\mathcal{F}$ is the set of feasible average rates.
- $\mathcal{T}$ is the set of feasible average transport rates.

Since $\mathbf{x}^n(s)$ has dimension L (where L is a number of links), by virtue of the Carathéodory theorem, it is enough to consider $N = L + 1$ time slots of arbitrary lengths $\{\boldsymbol{\alpha}_n(s)\}_{n=1\cdots N}$ in order to achieve any point in the convex closure of points $\mathbf{x}^n(s)$.

We are interested in the set $\mathcal{F}$ of feasible long-term average flow rates, which is the average over the system state S, and over slots allocations in each system state. It is the set of $\bar{\mathbf{f}} \in \mathbb{R}^I$ such that for each system state $s \in \mathcal{S}$ there exist a schedule $\boldsymbol{\alpha}_n(s)$, a set of power allocations $\mathbf{p}^n(s)$, corresponding set of rate allocations $\mathbf{x}^n(s)$ and average rates and powers $\bar{\mathbf{x}}(s)$ and $\bar{\mathbf{p}}(s)$, such that the following set of equalities and inequalities are satisfied:

$$\bar{\mathbf{f}} = \mathbb{E}\{\mathbf{f}(S)\} \tag{4.2}$$

$$\mathbf{f}(s) = F(s)\mathbf{y}(s) \tag{4.3}$$

$$\bar{\mathbf{x}}(s) \geq R(s)\mathbf{y}(s) \tag{4.4}$$

$$\bar{\mathbf{x}}(s) = \sum_{n=1}^{L+1} \alpha_n(s)\mathbf{x}^n(s) \tag{4.5}$$

$$\mathbf{x}_l^n(s) = r(\mathbf{SINR}_l^n(s)) \tag{4.6}$$

$$\mathbf{SINR}_l^n(s) = \frac{\mathbf{p}_l^n(s)h_{ll}(s)}{\eta_l(s) + \sum_{k\neq l} \mathbf{p}_k^n(s)h_{kl}(s)} \tag{4.7}$$

$$\mathbf{p}_l^n(s) \leq P_l^{MAX} \tag{4.8}$$

$$\bar{\mathbf{p}}_l \leq \overline{P}_l^{MAX} \tag{4.9}$$

$$\bar{\mathbf{p}} = \mathbb{E}\{\sum_{n=1}^{L+1} \alpha_n(s)\mathbf{p}^n(s)\} \tag{4.10}$$

$$1 = \sum_{n=1}^{L+1} \alpha_n(s), \tag{4.11}$$

$$1 \geq \sum_{l:l.\mathrm{src}=o} 1_{\{p_l^n(s)>0\}} + \sum_{l:l.\mathrm{dst}=o} 1_{\{p_l^n(s)>0\}}, \tag{4.12}$$

where $l.\mathrm{src} = o$ and $l.\mathrm{dst} = o$ are true if node o is the source or the destination of link l, respectively. Function $1(x)$ denotes the identity function; it is 1 if x is true, or else it is 0.

Equation (4.6) and Equation (4.7) give SNRs and rates for all links in all slots and all states. The average rates of links during a schedule, and in a given state, are given by Equation (4.5). Routing for a given state is defined by Equation (4.3) and Equation (4.4). Equation (4.8), Equation (4.9) and Equation (4.10) define power constraints. Medium access constraints are given by Equation (4.12). The end-to-end rates, average with respect to the system states, are given in Equation (4.2).

We note that if $\bar{\mathbf{x}}(s) = R(s)\mathbf{y}(s)$ there might be buffering problems. However, as shown in [66], any point in the interior of the set $\mathcal{F}$ is feasible, and the system is stable, for a large class of traffic patterns and system state distributions, thus we can approach arbitrarily close to the equality in Equation (4.4). To avoid an additional complexity of the presentation, we will allow $\bar{\mathbf{x}}(s) = R(s)\mathbf{y}(s)$.

4.4.2 Feasible Set of Transport Rates

In [30], the transport rate of a flow is defined as the rate of a flow multiplied by the distance covered by the flow between the source and destinations (call this $\mathrm{len}(i)$ for flow i). Therefore, we define the set of feasible long-term average transport rates $\mathcal{T}$ as

$$\mathcal{T} = \{\mathbf{t} \in \mathbb{R}^I \mid (\exists \bar{\mathbf{f}} \in \mathcal{F})\ \mathbf{t}_i = \bar{\mathbf{f}}_i \mathrm{len}(i)\} \tag{4.13}$$

4.4.3 Design Criteria

Given network technology, for each topology and traffic demand there is a given set of feasible rates $\mathcal{F}$ and a set of transport rates $\mathcal{T}$. We consider optimizing the system according to one of the following criteria:

Rate Criteria

- **capacity**: maximize $\sum_{i=1}^{I} \bar{\mathbf{f}}_i$ over all $\bar{\mathbf{f}} \in \mathcal{F}$
- **max-min fairness**: find the max-min fair rate vector $\bar{\mathbf{f}}^*$ in $\mathcal{F}$
- **utility fairness**: maximize $\sum_{i=1}^{I} U(\bar{\mathbf{f}}_i, \xi)$ over all $\bar{\mathbf{f}} \in \mathcal{F}$

It is easy to verify that $\mathcal{F}$ and $\mathcal{T}$ are convex and compact. The first and the third criteria are defined by concave maximization problems over $\mathcal{F}$, thus they always have a solution. Also, as shown in Chapter 3, max-min fair allocation exists on all convex and compact sets, which is the case here. Thus, all the three metrics are always well defined.

The max-min fair allocation does not have $\mathbf{x}_i^* = \mathbf{x}_j^*$ in general for $i \neq j$, even on convex sets (see Section 4.5.6 for some examples).

In general, the rate vectors that satisfy each of the three criteria are significantly different, as illustrated by the examples in the following sections.

Transport Rate Criteria

Similarly, we define:

- **transport capacity**: maximize $\sum_{i=1}^{I} \bar{\mathbf{f}}_i \mathrm{len}(i)$ over all $\bar{\mathbf{f}} \in \mathcal{F}$
- **transport-max-min fairness**: find the max-min fair transport rate vector $\mathbf{t}^*$ in $\mathcal{T}$
- **transport-utility fairness**: maximize $\sum_{i=1}^{I} U(\bar{\mathbf{f}}_i \mathrm{len}(i), \xi)$ over all $\bar{\mathbf{f}} \in \mathcal{F}$

A nice feature of the proportional fairness criterion ($\xi = 1$) is that the transport proportional fairness and proportional fairness lead to the same objective (up to a constant) and need not be considered separately. In contrast, the rates that maximize utility for any $\xi \neq 1$ differ from the rates that maximize the corresponding transport utility. Existence and uniqueness hold for transport criteria in the same way as for rate criteria.

4.4.4 Performance Indices

In the rest of this chapter, we evaluate the properties of the optimal rates that correspond to each of the criteria above. It is convenient to use indices that quantify efficiency and fairness.

The **efficiency index** of a feasible rate $\bar{\mathbf{f}}$ in a given feasible set $\mathcal{F}$ is $\frac{\sum_{i=1}^{I} \bar{\mathbf{f}}_i}{\sum_{i=1}^{I} \bar{\mathbf{f}}_i^c}$, where $\bar{\mathbf{f}}^c$ is the rate vector that maximizes capacity in $\mathcal{F}$. It is always between 0 and 1.

Similarly, the **transport efficiency index** of $\bar{\mathbf{f}}$ in $\mathcal{F}$ is $\frac{\sum_{i=1}^{I} \bar{\mathbf{f}}_i \text{len}(i)}{\sum_{i=1}^{I} \bar{\mathbf{f}}_i^t \text{len}(i)}$, where $\bar{\mathbf{f}}^t$ is the rate vector that maximizes transport capacity in $\mathcal{F}$.

The **max-min fairness index** of $\bar{\mathbf{f}}$ in $\mathcal{F}$ is thus defined as $\frac{(\sum_i \bar{\mathbf{f}}_i^* \bar{\mathbf{f}}_i)^2}{(\sum_i \mathbf{f}_i^{*2})(\sum_i \mathbf{f}_i^2)}$, where $\bar{\mathbf{f}}^*$ is the max-min fair element of $\mathcal{F}$. The max-min fairness index is defined as $\cos^2 \alpha$, where α is the angular deviation from $\bar{\mathbf{f}}$ to the max-min fair allocation $\bar{\mathbf{f}}^*$ in $\mathcal{F}$. The max-min fairness index is between 0 and 1; it is equal to 1 if $\bar{\mathbf{f}}$ is proportional to the max-min fair allocation of rates. The smaller it is, the less fair the allocation is. When the number of flows I is large, the minimum value of the max-min fairness index is close to 0.

Our max-min fairness index coincides with Jain's definition of fairness index [41] in the case where the max-min fair allocation $\bar{\mathbf{f}}^*$ has all components equal. Otherwise, it differs.

Similarly, the **transport max-min fairness index** of $\bar{\mathbf{f}}$ in $\mathcal{F}$ is $\frac{(\sum_i \mathbf{t}_i^* \bar{\mathbf{f}}_i \text{len}(i))^2}{(\sum_i \mathbf{t}_i^{*2})(\sum_i (\bar{\mathbf{f}}_i \text{len}(i))^2)}$, where $\mathbf{t}^*$ is the max-min fair element of $\mathcal{T}$

4.4.5 Performance Metrics

The indices defined above require computing the reference rate vector that is optimal with respect to a design criterion, and they depend on the set of rate vectors that is being considered. In contrast, metrics are defined as a function of the rate alone, independent of any set of rate vectors. For completeness, we now give the metrics that correspond to the design criteria defined above. They may be useful in practical situations where, unlike in this work, the computation of the reference rates is not feasible. This occurs for example when a protocol is given by its implementation in a simulator and the feasible set is hard to define explicitly.

For a rate vector $\bar{\mathbf{f}}$, the **capacity metric** is $\sum_{i=1}^{I} \bar{\mathbf{f}}_i$ and the **transport capacity metric** is $\sum_{i=1}^{I} \bar{\mathbf{f}}_i \text{len}(i)$. They both measure the efficiency of $\bar{\mathbf{f}}$.

A metric that corresponds to max-min fairness is more difficult to define. Many authors use Jain's fairness index defined above, but this is not always appropriate. Indeed, it measures the deviation from an ideal rate vector where all components are equal, and this is not necessarily the fairest vector. A more accurate, but more complex, metric uses leximin ordering [54, 82]. It is not a real number in the usual sense. Instead, the **fairness metric** $f(\bar{\mathbf{f}})$ of a rate vector $\bar{\mathbf{f}}$ is the list of all its components in increasing order, and we say that a rate vector $\bar{\mathbf{f}}^1$ is fairer than a rate vector $\bar{\mathbf{f}}^2$ if $f(\bar{\mathbf{f}}^1)$ is larger than $f(\bar{\mathbf{f}}^2)$ in lexicographic order. The max-min fair vector is the fairest, in the sense of this metric. Similarly, the **transport fairness metric** is defined as the order statistic of the vector of transport rates $(\bar{\mathbf{f}}_i \text{len}(i))_i$.

The ξ-**utility** is $\sum_{i=1}^{I} U(\bar{\mathbf{f}}_i, \xi)$. This defines a family of metrics, obtained through a different choice of parameter ξ. One example is logarithmic utility, corresponding to the proportional

fairness. The family also includes capacity metric ($\xi = 0$). However, the metric is not defined for $\xi \to \infty$ and we cannot use this approach to derive a metric for max-min fairness.

Analog to that, we defined the ξ-**transport utility metric** as $\sum_{i=1}^{I} U(\bar{\mathbf{f}}_i \text{len}(i), \xi)$. As noted in the discussion about performance objectives, the ξ-transport utility and ξ-utility of a rate allocation differ in general, except for $\xi = 1$ where they differ by a constant additive factor, thus they can be regarded as the same metric.

By using the utility performance metric, one can assess both efficiency and fairness. The importance of one or another is controlled by varying parameter ξ. We show that both $\xi = 0$ and very large ξ are not appropriate since they completely ignore fairness and efficiency, respectively. We also show that the design criteria based on values of ξ around 1 are the best, in the sense of robustness, against efficiency or fairness anomalies, for various power limitations. This suggests using logarithmic utilities ($\xi = 1$) as a metric of choice for evaluating ad-hoc wireless networks.

4.5 Max-min Fairness

In this section, we analyze properties of the max-min fair allocation. We show that there exists a class of convex sets with the property that a max-min fair vector on such a set has all components equal. We then show that a set of feasible long-term average rates in any wireless network without long-term average power constraints, modeled by Equations (4.2) - (4.12), admits this property thus implying that the rates in max-min fair allocation have to be equal.

4.5.1 Solidarity Property and Equality

Let us consider a class of sets in $\mathbb{R}^n$ with a property that for any feasible point we can trade a sufficiently small value of one component for a sufficiently small value of other components. More precisely, we define the solidarity property as follows:

Definition 4.1 *A subset $\mathcal{X}$ of $\mathbb{R}^n$ has the solidarity property iff for all $i, j, i \neq j$, for all $\mathbf{x} \in \mathcal{X}$ such that $\mathbf{x}_i > 0$, and for all $\epsilon > 0$ small enough, there exist positive $0 \leq \alpha_i < \epsilon$, $0 < \alpha_j < \epsilon$ and for all $k \neq i, k \neq j$, there exists $0 \leq \alpha_k < \epsilon$ such that $\mathbf{y} = \mathbf{x} - \alpha_i \mathbf{e}_i + \alpha_j \mathbf{e}_j + \sum_{k \neq i, k \neq j} \alpha_k \mathbf{e}_k$ belongs to $\mathcal{X}$.*

Not all sets have the solidarity property. In particular, not all convex sets have the solidarity property. Simple examples of networks with feasible rate sets with and without the solidarity property are given in Figure 4.2.

Figure 4.2: On the left, there is an example of a wireless network whose set of feasible rates has the solidarity property. Rate of flow 12 is constrained by $f_{12} \leq r(P_{12}h_{12}/(N + P_{34}h_{32}))$. Rate of flow 34 is constrained by $f_{34} \leq r(P_{34}h_{34}/(N + P_{12}h_{14}))$. It is always possible to increase the rate of one flow at the expense of the other. On the right, there is an example of a wired network whose feasible rate set does not have the solidarity property. Flows f_{12} and f_{13} are constrained by $f_{12} + f_{13} \leq x_{12}$ and $f_{13} \leq x_{23}$. When flow 13 hits limit on link 23 it cannot be further increased by decreasing rate of flow 12.

A characteristic of a set with the solidarity property is that all components of the max-min fair vector are equal. This is formulated in the following proposition:

Proposition 4.1 *If a set $\mathcal{X}$ has the solidarity property, then the max-min fair allocation $\mathbf{x}$ on $\mathcal{X}$ has all components equal: $\mathbf{x}_i = \mathbf{x}_j$ for all i, j, if the max-min fair allocation on $\mathcal{X}$ exists.*

Proof: Let us denote by $\mathbf{x}$ a max-min fair allocation on $\mathcal{X}$ and let us assume the contrary, that there exists i and j such that $\mathbf{x}_i - \mathbf{x}_j > 2\epsilon$ for some $\epsilon > 0$. Then, according to the solidarity property, there exists $\mathbf{y}$ such that $\mathbf{x}_i \geq \mathbf{y}_i > \mathbf{x}_i - \epsilon > \mathbf{x}_j + \epsilon > \mathbf{y}_j > \mathbf{x}_j$, and $\mathbf{y}_k \geq \mathbf{x}_k$ for all $k \neq i, k \neq j$ which contradicts with the definition of max-min fairness.

q.e.d.

4.5.2 Solidarity of The Feasible Rate Set of A Multi-hop Wireless Network

The feasible set of a wired network is given with a set of linear constraints. It is convex, but in general it does not have the solidarity property, as can be seen on the right of Figure 4.2. In the case of an ad-hoc wireless network, defined under the framework from Section 4.4, we show that the feasible rate set of any such network without long-term average power constraints, has the solidarity property.

Proposition 4.2 *Any feasible rate set $\mathcal{F}$ given by a set of equalities and inequalities (4.2)-(4.12), assuming $\overline{P}_l^{MAX} > P_l^{MAX}$ for all links l, has the solidarity property. Also, a feasible transport rate set $\mathcal{T}$ given by Equation (4.13) has the solidarity property.*

Proof: Let us denote by $(\mathbf{y}(s), \bar{\mathbf{x}}(s), \boldsymbol{\alpha}(s), (\mathbf{p}^i(s))_{n=1\cdots N})_{s\in\mathcal{S}}$ the values of slack variables, used in the constraint set given by eq. (4.2)-(4.12), that satisfies these constraints for rates $\bar{\mathbf{f}}$. We want to show that for every i and j, such that $\bar{\mathbf{f}}_i > 0$, it is possible to increase $\bar{\mathbf{f}}_j$ by an arbitrary small value, by decreasing $\bar{\mathbf{f}}_i$.

We first show that the average power limit is always satisfied if $P_l^{MAX} \leq \overline{P}^{MAX}$. We have that for every link l, slot n and state s, the transmission power is limited by $p_l^n(s) \leq P_l^{MAX}$. We then have $\bar{\mathbf{p}}_l \leq P_l^{MAX} \leq \overline{P}^{MAX}$, and we conclude that further on we do not have to explicitly consider the average power constraint.

Since $\bar{\mathbf{f}}_i > 0$, there exist a nonempty set $\mathcal{S}^+ \subseteq \mathcal{S}$ such that for all $s \in \mathcal{S}^+$ we have $\mathbf{f}_i(s) > 0$. We will first show by contradiction that for every $s \in \mathcal{S}^+$ we can increase $\mathbf{f}_j(s)$ by decreasing $\mathbf{f}_i(s)$. Then we will show that this will in turn increase $\bar{\mathbf{f}}_j$.

We first choose an arbitrary $s \in \mathcal{S}^+$, and we proceed by contradiction. Consider a feasible rate $\mathbf{f}(s)$ such that we cannot increase $\mathbf{f}_j(s)$ by decreasing $\mathbf{f}_i(s) > 0$.

For every state $s \in \mathcal{S}$ there exist a set $K(s,j) \neq \emptyset$ of links (rows in matrice $R(s)$) such that, for each $k \in K(s,j)$, there exists path m with positive rate ($y_m(s) > 0$), that belongs to flow j ($F_{mj}(s) = 1$), and passes over link k ($R_{km}(s) = 1$), and there is a strict equality $\bar{\mathbf{x}}_k(s) = (R(s)\mathbf{y}(s))_k$ (else, we can increase $\mathbf{f}_j(s)$, at no cost). In plain words, $K(s,j)$ is a nonempty set of bottleneck links for flow j.

Suppose that for each link k from $K(s,j)$ there exist path m that has a positive rate ($y_m(s) > 0$), belonging to flow i, that passes over link k. If we decrease rates of all such paths by some ϵ, we decreased $\mathbf{f}_i(s)$, and in the newly obtained rate allocation we have for all k, $\bar{\mathbf{x}}_k(s) > (R(s)\mathbf{y}(s))_k$. This in turn means that we can increase $\mathbf{f}_j(s)$ that leads to contradiction. Therefore, we can find a link l that is not a bottleneck for flow j ($l \notin K(s,j)$) and on which flow i has a positive throughput (there exists path m such that $y_m(s) > 0, F_{mi}(s) = 1, R_{km}(s) = 1$).

Let us denote with $\mathrm{Com}(k) = \{m \in 1\cdots L \,:\, k.\mathrm{src} = m.\mathrm{src} \vee k.\mathrm{src} = m.\mathrm{dst} \vee k.\mathrm{dst} = m.\mathrm{src} \vee k.\mathrm{dst} = m.\mathrm{dst}\}$ a set of links that cannot be scheduled at the same time as link k since they share a common node. We pick an arbitrary $k \in K(s,j)$ (we have $l \neq k$), a slot n when link l is active, and divide it in three slots, n_1, n_2 and n_3 of lengths $\alpha_{n_1}(s), \alpha_{n_2}(s), \alpha_{n_3}(s) > 0$ respectively, such that $\alpha_{n_1}(s) + \alpha_{n_2}(s) + \alpha_{n_3}(s) = \alpha_n(s)$. In the first slot we keep the same power allocation as in slot n. In the second slot we turn off the link l, and leave the rest as it is. In the third slot we turn off link l and all of the links from $\mathrm{Com}(k)$, and increase, if necessary, the power of link k such that $0 < p_k^{n_3}(s) \leq P_k^{MAX}$ and the interferences perceived by other active users is smaller than in the original scheduling of slot n. As we have show above, this is always possible since $\overline{P}_k^{MAX} \geq P_k^{MAX}$ for all k, and all links from $\mathrm{Com}(k)$ are silent.

In the new scheduling, links belonging to $\mathrm{Com}(k)$ have the same rates in slot n_1, higher rates in slot n_2 and lower rates in slot n_3. It is easy to see that there exists a small $\epsilon_k > 0$ and a choice of $\alpha_{n_1}(s), \alpha_{n_2}(s), \alpha_{n_3}(s)$ such that all links have the same or higher rates, except for link k whose rate has decreased by at most ϵ^l and link k whose rate has increased by at least ϵ^k. We repeat the same for all $k \in K(s,j)$.

We thus have a new average link rate allocation $\bar{\mathbf{x}}'(s)$ such that $\bar{\mathbf{x}}_l(s) - \sum_{k\in K(s,i)} \epsilon_k < \bar{\mathbf{x}}'_l(s) < \bar{\mathbf{x}}_l(s)$ and $\bar{\mathbf{x}}_k(s) < \bar{\mathbf{x}}'_k(s) < \bar{\mathbf{x}}_k(s) + \epsilon_k$, for all $k \in K(s,i)$. Now we can increase $\mathbf{f}_j(s)$ by at least $\min_{k\in K(s,j)} \epsilon_k$, by decreasing $\mathbf{f}_i(s)$ by at most $\sum_{k\in K(s,i)} \epsilon_k$.

This leads to a contradiction, and we have proved that for every $s \in \mathcal{S}^+$ we can increase $\mathbf{f}_j(s)$ by some positive $\epsilon(s)$.

We repeat the same for every $s \in \mathcal{S}^+$. We then have the new average rate of flow j,

$$\begin{aligned}
\bar{\mathbf{f}}'_j &= \sum_{s\in\mathcal{S}^+} \mathbb{P}\{S=s\}\,(\mathbf{f}_j(s) + \epsilon(s)) \\
&= \bar{\mathbf{f}}_j + \sum_{s\in\mathcal{S}^+} \mathbb{P}\{S=s\}\,\epsilon(s), \\
&> \bar{\mathbf{f}}_j,
\end{aligned}$$

since $\sum_{s\in\mathcal{S}^+} P\{S=s\}\epsilon(s) > 0$. This proves the solidarity property of set $\mathcal{F}$.

The same reasoning holds for a set of transport rate, hence the second part of the statement.

q.e.d.

4.5.3 Equality of Max-min Fair Rates

Consider an arbitrary network where long-term average power constraints are larger than transmission power constraints. It is easy to verify that the feasible set given by constraints (4.2)-(4.12) is convex, hence according to Chapter 3 it has the max-min fair allocation. Since this set also has the solidarity property, we have the following:

Corollary 4.1 *The max-min fair average rate allocation of any network given by constraints (4.2)-(4.12), with no long-term average power constraints ($\overline{P}^{MAX} \geq P^{MAX}$), has all rates equal. The max-min fair transport rate allocation has all transport rates equal.*

Equality of rates implies that all flows, including the most inefficient ones, have an equal rate. This can be very inefficient in a heterogeneous network. For example, if one node is almost disconnected, then it will receive a rate close to zero. According to Corollary 4.1, all other flows will have the same rate.

Another example is given in Figure 4.3. On the left, we show an example of a network where 12 nodes (6 flows) are randomly placed on a square of 100m x 100m. The source and the destination of each flow are joined with a line. Each flow can use either the direct route or the minimum energy route. In this example, we assume no random fading, and we set all transmission power constraints to be equal to $P^{MAX}/\eta = 90\text{dB}$, where η is a white background noise. The actual SNR on each receiver depends on the distance between the source of the link and the destination of the link. For example, according to the UWB indoor path loss model [38], if a source sends to a destination that is 10 m away with maximum power and $P^{MAX}/\eta = 90\text{dB}$, we have SNR at the receiver around 10 dB. This in turn leads to the rate of 100 Mb/s within the framework of [21, 96].

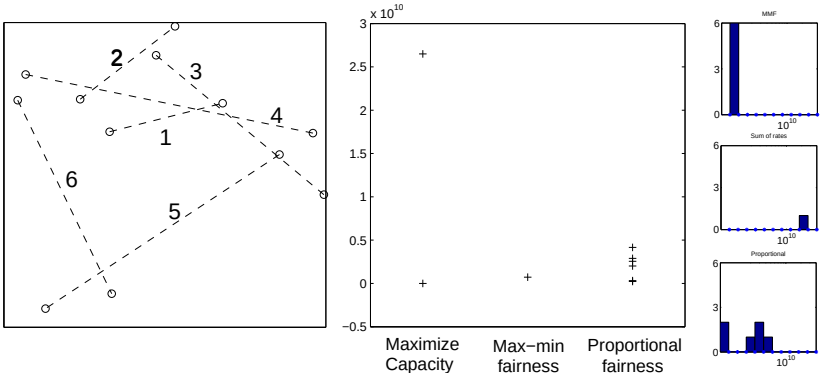


Figure 4.3: Left: example of random network topology. 12 nodes are randomly placed on a 100m x 100m grid. Nodes are depicted with circles, and source-destination pairs are joined with lines. Each flow uses the optimal routing (the direct or the minimum-energy route). Center: corresponding rate distribution (total capacity , max-min fairness and proportionally fairness). Right: Histograms of the three allocations.

In the middle and on the right of Figure 4.3, we see the optimal rate allocations with respect to the three metrics, for this example. We see that when maximizing total capacity , only flow 1, which is the shortest flow, has a positive rate, and the rates of other flows are zero. In the case of max-min fairness, all rates are the same. Proportional fairness exhibits larger variation in rates than max-min fairness, but it does not starve the least efficient flows. But, it is more efficient than max-min fairness. We also illustrate the Corollary 4.1 on more random examples on Figure 4.4 in Section 4.6.2.

From Corollary 4.1 we also see that in the case of the max-min fair transport rate allocation, all transport rates are equal. Obviously, the rates themselves are not equal as the flow lengths differ. Still in this case, as can be seen in the numerical examples from Figure 4.4 in Section 4.6.2, the corresponding rate allocation suffers from the same inefficiency problem.

4.5.4 Influence of Long-Term Average Power Constraint

Corollary 4.1 holds when there are no long-term average power constraints (which is equal to having long-term average power constraints greater or equal to transmission power constraints). When long-term average power constraints are smaller than the transmission power constraints, the max-min fair rates are not equal anymore. However, we see that for high transmission power constraints ($P^{MAX} \geq 40\text{dB}$, see Figure 4.4 in Section 4.6.2) and high long-term average power constraints ($\overline{P}^{MAX}/P^{MAX} \geq 0.5$, see Figure 4.8, Section Section 4.8) the max-min fair rate allocation is still inefficient.

On Figure 4.8 we also see that for very small long-term average power constraints, the optimal allocation becomes independent of the choice of the metric.

Overall, these arguments show that max-min fairness is not an appropriate metric even when long-term average power constraints exist.

4.5.5 An Application to an 802.11 Network

An example of the above findings can be seen in [7]. Consider an 802.11 network where several nodes send data directly to a single destination (base-station). Assume node 1 is far away and it codes for 1Mb/s, and the others are close enough to code for 11Mb/s. One would expect that node 1 achieves a smaller rate than the other nodes. However, as shown in [7], this is not the case, and all nodes achieve an effective throughput of around 1Mb/s.

According to the analysis done in [7], when a node gets an access to the network, it sends a packet of a fixed size, thus the occupancy time is inversely proportional to the coding rate. In other words, a node sends the same amount of data during a channel use, regardless of its coding rate. Let us consider a discrete random process X_t representing a user that occupies a channel during the t-th channel use. According to eq. (7) and (8) from [7], X_t is an i.i.d. uniform random process, and all nodes have an equal probability to obtain network access when the network is idle. This leads us straightforward to the following proposition:

Proposition 4.3 *An 802.11 network in DCF mode where all nodes talk directly to a single destination (hence there is no hidden terminal problem), implements max-min fairness*

In other words, the equality of rates observed in [7] is not solely a property of the 802.11 physical layer but rather of max-min fairness that is obtained in this specific example. This means that any other protocol that would implement max-min fairness, and would fit in the framework of eq. (4.2)-(4.12), would have the same inefficiency problem.

4.5.6 When Max-min Fairness Does Not Lead To Equality

We note that the assumptions of Corollary 4.1 are not true in general for any convex set, but only for those that have the solidarity property. To illustrate this, we give a few counter examples:

- *Wired Networks:* The corollary does not hold for a class of wired networks. For an example, see Figure 4.2 on the right.
- *Clustered Networks:* The corollary does not hold for a clustered wireless network. Assume a simple network of two links, link $(1,2)$ and link $(3,4)$, and assume it is clustered such that nodes 3 and 4 does not hear node 1 and 2 and vice versa (meaning that $h_{13} = h_{14} = h_{23} = h_{24} = 0$). Then, rates $\bar{f}_{12}$ and $\bar{f}_{34}$ are not going to be equal.
- *Long-Term Average Power Constraint:* The corollary does not hold if the long-term average power constraint is smaller than transmission power constraints, as shown in Section 4.8.

4.6 Maximizing Total Capacity

4.6.1 Asymptotic Results

As discussed previously, maximizing the total capacity metric is efficient but may lead to gross unfairness, especially in the case of large transmission power constraints. In order to demonstrate this, we first look at the asymptotic case. We start by considering a network without variable noise power and channel conditions (that is, with only one state), and we show that total capacity metric becomes very unfair as transmission power tends to infinity.

At this point, we need an additional assumption on the rate function, $\lim_{\mathrm{SNR}\to\infty} r(\mathrm{SNR}) = \infty$, that is we can increase the rate of a link arbitrarily high by sufficiently increasing the signal-to-interference-and-noise ratio on this link. We also assume here no long-term average power constraint, hence $\overline{P}^{MAX} \geq P^{MAX}$.

In order to simplify the presentation, we assume that all transmission power constraints are the same, that is for all l, $P^{MAX} = P_l^{MAX}$. This can be generalized for non-uniform power constraints, assuming that when P^{MAX} goes to infinity there exist fixed positive numbers γ_l such that $P_l^{MAX}/P^{MAX} > \gamma_l$.

Proposition 4.4 *Consider an arbitrary network with one system state ($|\mathcal{S}| = 1$), and let $\bar{\mathbf{f}}$ be the long-term average rate allocations that maximizes total capacity . Assume that when the*

signal-to-interference-and-noise ratio at a receiver $\mathbf{SINR}_i$ *tends to infinity the rate of link* i*,* $r(\mathbf{SINR}_i)$ *also tends to infinity, and assume there is no long-term average power constraint (or equivalently,* $\overline{P}^{MAX} \geq P^{MAX}$*). In a limiting case when* $P^{MAX} \to \infty$*, there will be one or more flows that have the same rate* $f = \Omega(r(P^{MAX}/\eta))$ *and all the others will have a rate that is* $o(r(P^{MAX}/\eta))$*. The same happens when considering transport rates.*

Since in the theorem we consider a network with only one system state, we will omit the state S from the notation throughout the proof.

We first propose a lemma that characterizes the optimal schedule and power allocation when transmission power limit tends to infinity.

Lemma 4.1 *Let* $\mathbf{p}^n = \mathbf{p}^n(P^{MAX})$ *be the optimal power allocation in slot* n *that maximizes the sum of rates for a given transmission power limit* P^{MAX}*. For all slots* n *there exists a link* i *such that both are true:*

1. *There exists* $\epsilon_i > 0$ *such that* $\limsup_{P^{MAX}\to\infty} \frac{\mathbf{p}_i^n}{P^{MAX}} > \epsilon_i$.
2. *For all* $j \neq i$, $\limsup_{P^{MAX}\to\infty} \frac{\mathbf{p}_j^n}{\mathbf{p}_i^n} = 0$.

Proof: The first statement can be interpreted as the following: there exists $\epsilon_i > 0$ such that for all Ω_i there exists $P^{MAX} > \Omega_i$ such that $\mathbf{p}_i^n/P^{MAX} > \epsilon_i$. We show the statement using contradiction. Suppose that for some slot n and for each link i and all $\epsilon_i > 0$ there exists Ω_i such that for all $P^{MAX} > \Omega_i$ we have $\mathbf{p}_i^n/P^{MAX} < \epsilon_i$. Let us choose an arbitrary link j and increase its power allocation in slot n to $\mathbf{p}_j'^n = P^{MAX}$. We then have the following

$$\frac{\mathbf{SINR}_j'^n}{\mathbf{SINR}_i^n} = \frac{P^{MAX}h_{jj}}{\mathbf{p}_i^n h_{ii}} \frac{\eta + \sum_{k\neq i} \mathbf{p}_k^n h_{ki}}{\eta + \sum_{k\neq j} \mathbf{p}_k^n h_{kj}}.$$

We first look at the second multiplier. Let us denote with $\hat{p}^n = \max_i p_i^n$. We can then write $p_i^n = C_i \hat{p}^n$, where $C_i \leq 1$, and for at least one i we have $C_i = 1$. Suppose first that $\hat{p}^n \leq \eta$. We than have

$$\frac{\eta + \sum_{k\neq i} \mathbf{p}_k^n h_{ki}}{\eta + \sum_{k\neq j} \mathbf{p}_k^n h_{kj}} \geq \frac{\eta}{\eta\left(1 + \sum_{k\neq j} C_k h_{kj}\right)} \geq \frac{1}{1 + \sum_{k\neq j} h_{kj}}.$$

On the contrary, if $\hat{p}^n > \eta$ we have

$$\frac{\eta + \sum_{k\neq i} \mathbf{p}_k^n h_{ki}}{\eta + \sum_{k\neq j} \mathbf{p}_k^n h_{kj}} \geq \frac{\sum_{k\neq i} C_k h_{ki}}{\frac{\eta}{\hat{p}^n} + \sum_{k\neq j} C_k h_{kj}} \geq \frac{\min_{k\neq i} h_{ki}}{1 + \sum_{k\neq j} C_k h_{kj}},$$

since at least for one k we have $C_k = 1$. Therefore, we can lowerbound

$$\frac{\eta + \sum_{k \neq i} \mathbf{p}_k^n h_{ki}}{\eta + \sum_{k \neq j} \mathbf{p}_k^n h_{kj}} \geq \frac{\min\{1, \min_{k \neq i} h_{ki}\}}{1 + \sum_{k \neq j} C_k h_{kj}} = K,$$

where K is a constant that depends only on the topology and not on P^{MAX}.

We now have

$$\frac{\mathbf{SINR}_j'^n}{\mathbf{SINR}_i^n} = \frac{P^{MAX} h_{jj}}{\mathbf{p}_i^n h_{ii}} \frac{\eta + \sum_{k \neq i} \mathbf{p}_k^n h_{ki}}{\eta + \sum_{k \neq j} \mathbf{p}_k^n h_{kj}} \tag{4.14}$$

$$\geq \frac{P^{MAX}}{\mathbf{p}_i^n} K > \frac{K}{\epsilon_i}. \tag{4.15}$$

Therefore, we can make new $\mathbf{SINR}_j'^n$ arbitrary higher than any signal-to-interference-and-noise ratio $\mathbf{SINR}_i^n$ in slot n. Due to the assumption on the rate function, in the same way we can make a rate of link j in slot n arbitrary larger than the original rates of other links in slot n, as well as the sum of rates of all links in slot n. In particular, if link j connects a source and a destination of a flow, by increasing $\mathbf{p}_j^n$ to P^{MAX} we increased the total rate, which contradicts with the initial assumption.

We next show the second statement. It is equivallent to saying that for all $j \neq i$ and for all $\epsilon_j > 0$ there exists Ω_j such that for all $P^{MAX} > \Omega_j$ we have $\mathbf{p}_j^n / \mathbf{p}_i^n < \epsilon_j$. Again, we use contradiction, and we suppose that in some slot n there exists link j such that for some ϵ_j and for all Ω_j there exists P^{MAX} such that $\mathbf{p}_j^n / \mathbf{p}_i^n > \epsilon_j$. Again, we consider a new power allocation where for some link l we set $\mathbf{p}_l'^n = P^{MAX}$ and all the other powers are zero. We have the following

$$\frac{\mathbf{SINR}_l'^n}{\mathbf{SINR}_i^n} = \frac{P^{MAX} h_{ll}}{\mathbf{p}_i^n h_{ii}} \frac{\eta + \mathbf{p}_j^n h_{ji} + \sum_{k \neq i,j} \mathbf{p}_k^n h_{ki}}{\eta}$$

$$> \frac{P^{MAX} \mathbf{p}_j^n}{\mathbf{p}_i^n} K > P^{MAX} \epsilon_j K.$$

Here K and ϵ_j are fixed constants and for an arbitrary Ω_j there exists $P^{MAX} > \Omega_j$ that satisfies the above inequality. This in turn means that we can make $\mathbf{SINR}_l'^n$ arbitrary larger than $\mathbf{SINR}_i^n$. The same applies for $\mathbf{SINR}_j^n$. We can do similarly for a link $k \neq i, k \neq j$ by virtue of Equation (4.15). As we shown above, if l is a link between a source and a destination of a flow, the new allocation increases total rate which contradicts with the initial assumption.

q.e.d.

Intuitively, Lemma 4.1 shows that in the optimal power allocation for very large power constraints, there should be only one link with a very large power active at a time. All the other links should be allocated very small powers. We proceed to the proof of the proposition.

Proof of Proposition 4.4: Let us denote with $(\mathbf{y}(s), \bar{\mathbf{x}}(s), \boldsymbol{\alpha}(s), (\mathbf{p}^i(s))_{n=1\cdots N})$ the optimal solution to optimization problem (4.2)-(4.12). Consider link i. From Equations (4.2)-(4.12) we have the following inequality $\sum_{p\ni i} y_p \leq \sum_n \alpha_n \mathbf{x}_i^n$. By Lemma 4.1 we know that in the optimal power allocation, in each slot there is exactly one link whose power is $\Omega(P^{MAX})$ and all other links have powers $o(P^{MAX})$.

Suppose link i achieves rates of order $\Omega(P^{MAX})$ in different slots, of total duration $\alpha_{n(i)}$ (if link i never achieves rates of order $\Omega(P^{MAX})$, then we have $\alpha_{n(i)} = 0$). In the rest of the schedule it achieves rates of order $o(P^{MAX})$. By ignoring rates of order $o(P^{MAX})$, we can group the rest of slots into one slot, called slot $n(i)$. The duration of this slot is $\alpha_{n(i)}$, and the rate achieved can be approximated with $\mathbf{x}_i^{n(i)} = \max_m \mathbf{x}_i^m$. We can then say $\sum_{p\ni i} y_p = \alpha_{n(i)} \mathbf{x}_i^{n(i)}$ (we can assume equality, since we otherwise can assign all extra time to other power allocations).

Suppose we have an additional time $\Delta\alpha$ to serve path p. We need to spread it on all links belonging to path p such that the rate of each link i increases by the same Δy_p. We need to have $\Delta y_p = \Delta\alpha / \left(\sum_{i\in p} \frac{1}{x_i^{n(i)}} \right)$. Now, since the total capacity is a sum of the rates on all paths, in order to maximize total capacity we will assign time only to links of those paths that have the highest increase factor $1/\left(\sum_{i\in p} \frac{1}{x_i^{n(i)}} \right)$, and will not serve the other paths letting them have zero rate.

The same happens in the case of transport rates, since the increase factor is the same as above, multiplied by a length of the corresponding flow. Consequently, the corresponding rates will also tend to zero.

q.e.d.

Proposition 4.4 tells us that if a signal-to-interference-and-noise ratio is high enough, then only the most efficient flows are going to divide all the capacity of the medium, whereas all other flows will starve. An example is shown on Figure 4.3. In this example, all flows are single hop, and the most efficient flow is the shortest one, which is flow 1. Therefore all flows, except for flow 1, have zero rates. In the following, we illustrate that the same problem occurs for a large range of realistic signal-to-interference-and-noise settings.

The proposition will not hold for a network with variable channel fading or noise power. Consider a single-hop flow over a long link (that is, a link with high average signal attenuation). As suggested in [90], when maximizing total capacity it is always optimal to schedule the flow with

the best channel conditions. In most cases, the long flow will not be scheduled. However, it can still happen that, due to variability in the network medium, the long flow has the best channel in some slots and the long-term average rate of this flow is larger than zero. Nevertheless, in a very asymmetric multi-hop network, the probability of the flow, with a poor average performance, to have the best realization is very small. A typical example is a multi-hop network that has very short and very long links, and single-hop and multi-hop flows. In those cases, as Proposition 4.4 indicates, the rates of flows with poor average conditions will be close to zero, as is illustrated on Figure 4.6.

4.6.2 Numerical Results

In the above section we have seen that an increase in transmission power constraints will eventually lead to all but some flows having zero rates. It is not clear what the realistic values of the constraints for which this phenomenon occurs are. From [21] we see the phenomenon has been observed in a realistic network, and in this section we investigate in which transmission power region it occurs. Another issue we evaluate is the effect of variable channel fading and noise power on fairness of maximizing total capacity metric.

Simulation Setting

In order to analyze the behavior of the total capacity performance metric for a realistic power setting, we numerically evaluated it on random network topologies. We adapted the framework from [19], which assumes a rate is a linear function of the signal-to-interference-and-noise ratio at a receiver (this also corresponds to an UWB model from [21]). It is shown in [19] that the optimal power allocation strategy that maximizes total capacity is either to send with maximal power or not to send at all. Nevertheless, the optimization problem is still exponentially complex so it was not possible to run simulations for more than 12 nodes (6 flows) and with no random fading (the effect of random fading and noise power is evaluated latter in this section).

For each flow we consider a multi-path routing with a set of routes that comprise nodes that are on the shortest path between the source and the destination. This is a suboptimal set of routes since, in the case of high congestion in one area of a network, the optimal path may avoid that area even if it is not the shortest one. However, in most cases this heuristic is a good approximation, and it simplifies our calculation. Furthermore, by running tests on several random topologies, we concluded that in all cases the optimal heuristic among those is either the minimum energy route (relaying over intermediate nodes that minimizes total dissipated power), or the direct route (send

directly to the destination without relaying). Since constraining on these two routes for each flow further reduces the complexity of optimization, we used these heuristics to produce the results. In our example of networks with 12 nodes (6 flows), average number of hops per route is 2.19.

Uniform Topologies

We first consider uniform random network topologies with 12 nodes uniformly distributed on a square of 100m x 100m. Half of them are sources sending data, each to its own destination among the other half. All nodes are assumed to have the same transmission power constraints. We are looking for routing, scheduling, and power control that maximize the total capacity . An example of such a network, together with the optimal end-to-end rates with respect to different objectives, is given in Figure 4.3.

In Figure 4.4, on the top left, we show average fairness indices of the optimal rates with respect to total capacity and utility metrics, as well as the confidence intervals. On the x-axis, a ratio between maximal transmitting power and noise in dB is given.

From the numerical results depicted in Figure 4.4, on the top left, we see that maximizing total capacity leads to an acceptable fairness in the case of small transmission power limits. However, for large transmission power limits we see that maximizing total capacity exhibits high unfairness. The minimal value of fairness index is $1/I = 0.17$ and corresponds to the case when only one flow has a non-zero rate. The fairness index of total capacity drops to minimum for high power limits, as predicted by Proposition 4.4. We can see that this happens already for $P^{MAX}/\eta > 70dB$, and it happens for smaller power constraints when topology is non-uniform. These results confirm the unfairness observations made in [21], and show they are a consequence of the performance metrics rather than physical layer particularities. All these results are for unlimited battery lifetime constraints. However, the unfairness exists for a limited battery lifetime; for details, see Section 4.8.

Next, we used the three metrics to find the transport-optimal solutions on the set of rates $\mathcal{F}$. We calculated the transport fairness and the transport efficiency indices of the optimal rate allocations. This can be seen in Figure 4.4, on the right. We see that when considering the transport fairness index, we have similar results as for the fairness index: max-min fairness is the most transport-fair metric, and maximizing total capacity is the least. Similarly, maximizing total capacity is the most transport-efficient metric, and max-min fairness is the least. We also see a drop of the transport-efficiency of rate based metrics for $P^{MAX}/\eta = 40$dB. Thus, there is a major transport-inefficiency of the optimal rate allocations, but only in the case of intermediate transmission power limits. Intuitively, when powers are very small, there is little interference

and interactions between links. Hence maximizing capacity yields similar performances to maximizing transport capacity. When powers are high, whether maximizing capacity or transport capacity, radio resources are allocated to the most efficient link, and the two performances are similar again. For intermediate powers, the two metrics differ, and we see a large dip.

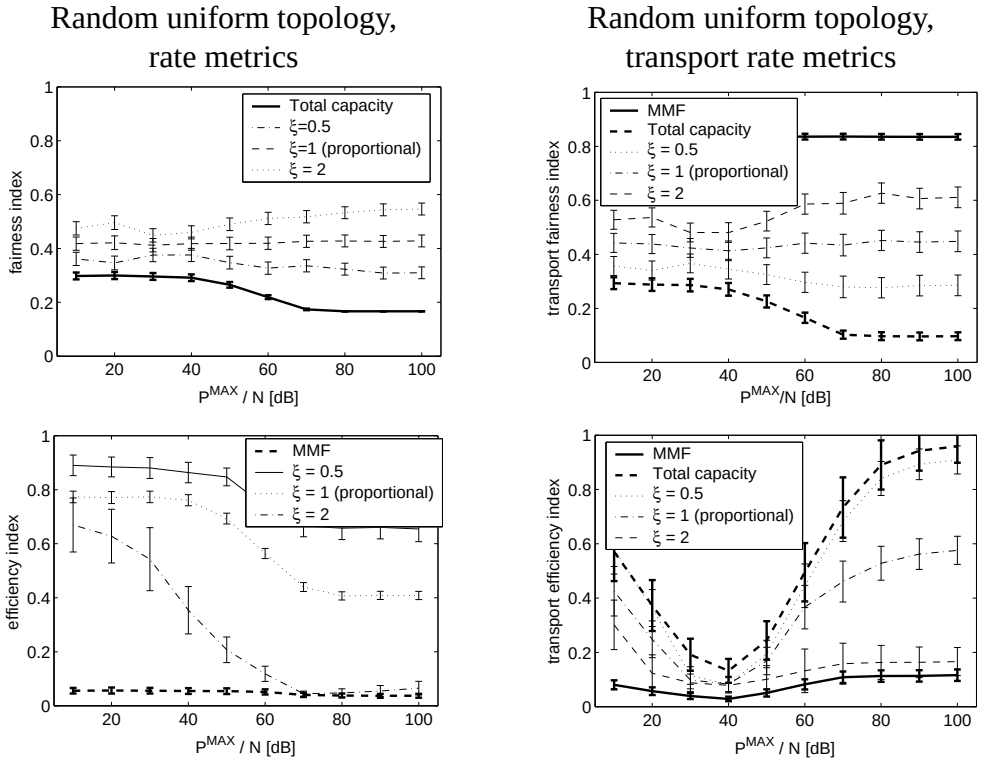


Figure 4.4: The fairness and efficiency indices of optimal rates with respect to different performance objectives ($\xi = 1$ means proportional fairness), versus the ratio of maximal transmitting power and noise: On the top, the fairness index is given; on the bottom is the efficiency one (note that both the fairness index of the max-min fair allocation and the efficiency index of the maximal total capacity are one). On the left, we see the indices for the three performance objectives applied to the set of feasible rates. On the right, we see the indices for the three performance objectives applied to the set of feasible transport rates. In both cases nodes are uniformly spread over the entire 100m x 100m square. Values of P^{MAX}/η on the x axis represent a realistic values that can be found on existing UWB or CDMA systems. In all cases we put no constraints on long-term average power . All figures show 95% confidence intervals. Notice that the minimal value of fairness index is $1/I$, where I is the number of flows; in this example, the smallest fairness index is 0.17 and represents that only one flow has a positive rate and all others have the zero rates.

Non-uniform Topologies

We also analyzed the fairness index of the optimal rates in a case of random non-uniform networks. We consider two non-uniform scenarios. The first one has a non-uniform node distribution but a uniform source-destination distribution. We again considered a square area of 100m x 100m and we divided it into 4 equal sub-squares of 50m x 50m each. We placed, in total, 12 uniformly distributed nodes in the upper left and lower right sub-squares. Each node chose uniformly one destination among all other nodes. We thus had several short and several long flows, and a hot-spot in the center of the big square. The results are depicted in Figure 4.5 on the left.

The second scenario has a uniform node distribution and a non-uniform source-destination distribution. A base station is placecd in the middle of a 100m x 100m square. Nodes are randomly distributed over the square, and they all talk to the base station. In this scenario we consider 6 flows, in order to maintain the same number of flows as in the other simulations. The results are depicted on the right in Figure 4.5.

Variable Channel Fading and Noise Power

Finally, we analyze the case with random fading and random noise powers. We assume that the fading between the source of link i and the destination of link j belongs to the set of possible fadings H_{ij}^s, and that the power of the white noise at the destination of link i belongs to the set of possible powers η_i^s. Then the set of possible states of the system is $\mathcal{S} = H_{11}^s \times \cdots \times H_{1L}^s \times \cdots \times H_{LL}^s \times \eta_1^s \times \cdots \times \eta_L^s$. This means that the number of possible system states $|S|$ grows exponentially with the number of links in the network and the number of possible states for each link. Since we have to find the optimal schedule for each system state, it is difficult to numerically evaluate this problem, even for small networks. In order to show the effect of randomness in a network, we consider a case of networks with two single-hop flows. We assume that the set of possible fadings is $H_{ij}^s = \{h_{ij}, 0.75h_{ij}, 0.5h_{ij}\}$ with probabilities $\{0.75, 0.2, 0.5\}$ respectively, where h_{ij} is the attenuation that depends on the distance, as used in the fixed fading case. We assume the same distribution for the noise power at receivers. We then construct the set of possible system states $\mathcal{S}$ with the corresponding probabilities. Using the same numerical approach as in the case of fixed fading, this time for every possible state, we calculate the optimal rates with respect to different metrics.

The results are depicted in Figure 4.6. Max-min fair rate allocation is as inefficient as it is in the case of constant fading and noise power. Total capacity is less unfair than it is in the case of constant fading. However, we still see that for high power limits ($P^{MAX}/\eta > 70dB$), one flow will have the zero rate (that is, fairness index will be equal to 0.5).

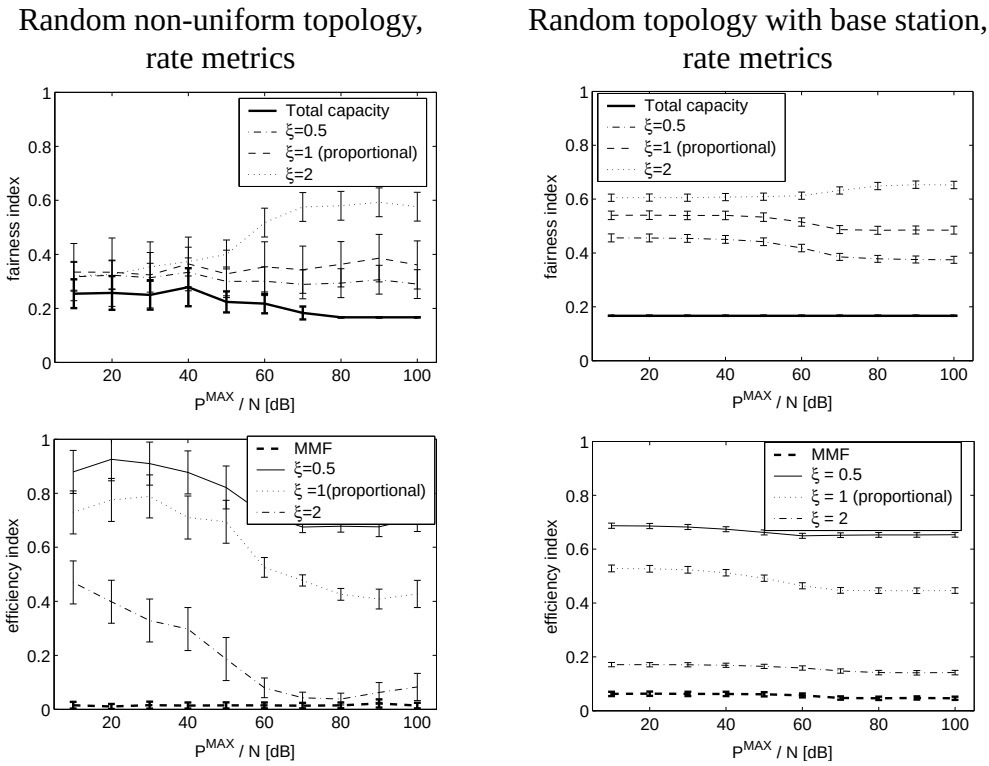


Figure 4.5: Non-uniform topologies - On the left we see the indices for the performance objectives ($\xi = 1$ means proportional fairness) applied to the set of feasible rates in the case when nodes are distributed only in the upper-left and lower-right quarters of a 100m x 100m square: On the top, the fairness index is given; on the bottom is the efficiency one (note that both the fairness index of the max-min fair allocation and the efficiency index of the maximal total capacity are one). On the right we see the indices for the case when nodes are uniformly distributed and they all talk to a base station located in the center of the square In all cases we put no constraints on long-term average power . All figures show 95% confidence intervals. Notice that the minimal value of fairness index is $1/I$, where I is the number of flows; in this example, the smallest fairness index is 0.17 and represents that only one flow has a positive rate and all others have the zero rates.

4.7 Utility Fairness

As seen in the previous sections (e.g. Figure 4.3), both maximizing total capacity and max-min fairness suffer from either inefficiency or unfairness. In this section, we analyze utility fairnesses in detail. We numerically evaluated the efficiency and fairness of the proportional fairness metric using the same setting as in Section 4.6.2. We show that proportional fairness ($\xi = 1$) in particular represents a robust compromise between efficiency and fairness.

As explained in Section 4.6, the optimal power allocation strategy that maximizes total capacity is either to send with maximal power or not to send at all. It is shown in [73], that, within

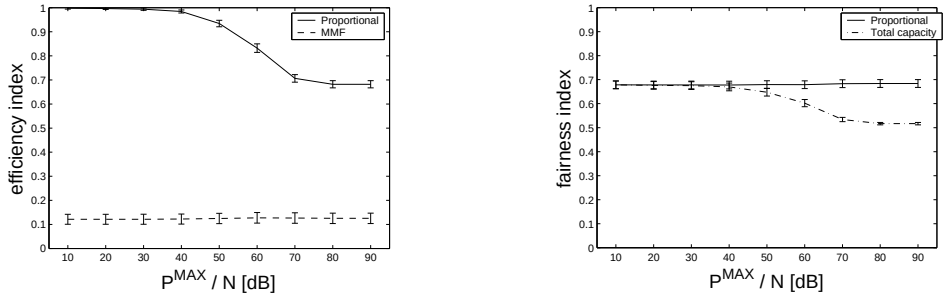


Figure 4.6: Random fading and noise power: Efficiency index (left) and fairness index (right) of maximum total capacity , max-min fair and the proportionally fair ($\xi = 1$) rates. Results are obtained on random uniform network topologies with 4 nodes (2 flows). Both fading and noise power take a fraction of $\{1, 0.75, 0.5\}$ of the constant fading case, with probabilities $\{0.75, 0.2, 0.5\}$, respectively. Note that in this example, fairness index of 0.5 means the one flow has zero rate.

the same framework, the same findings hold in the case of utility fairness. Therefore, we use the same numerical approach as explained in Section 4.6, to calculate different utility fairnesses. We also use the same routing heuristics.

The fairness and the efficiency indices of utility fair rate allocations for different values of ξ are depicted in Figure 4.4 on the top and bottom left, respectively. We give the results for frequently used proportional fairness ($\xi = 1$) and minimum potential delay ($\xi = 2$), and also for $\xi = 0.5$. We see that the difference between the design objectives are smaller when power limits are small (except for max-min fairness). This is due to the fact that the interference, hence the mutual interactions among nodes transmitting in parallel, become smaller when power is small.

We next analyze the case of large power limits ($P^{MAX}/\eta > 70$dB). For $\xi = 0.5$, the fairness index of the optimal rate allocation is close to the fairness index of the one that maximizes sum of rates, hence it exhibits the same unfairness problems. For $\xi = 2$, the efficiency index of the optimal rate allocation becomes very close to the one of max-min fair rate allocation, hence the optimal allocation is inefficient. The proportionally fair rate allocation ($\xi = 1$) is the most robust. Its efficiency remains constant for all values of transmission power constraint, and it is up to 10 times more efficient than the max-min fair allocation for high transmission power constraints.

We analyzed the efficiency index of the optimal rates in a case of random non-uniform networks, as above, and the results are depicted in Figure 4.5. On the left we considered a non-uniform node distribution, and on the right we considered a non-uniform traffic distribution. The fairness indices are given in the top row and the efficiency indices are given in the bottom row. The results are similar to those from the symmetric case, and the same conclusions hold.

On the right of Figure 4.4 we depict the transport fairness and transport efficiency properties

of the optimal rates on set $\mathcal{F}$. Max-min fairness is again much less transport-efficient than utility fairnesses and maximizing total capacity is much less transport fair than utility fairnesses. Like in the case of fairness and efficiency indices, proportional fairness ($\xi = 1$) is the most robust, and balanced with respect to transport-efficiency and transport-fairness.

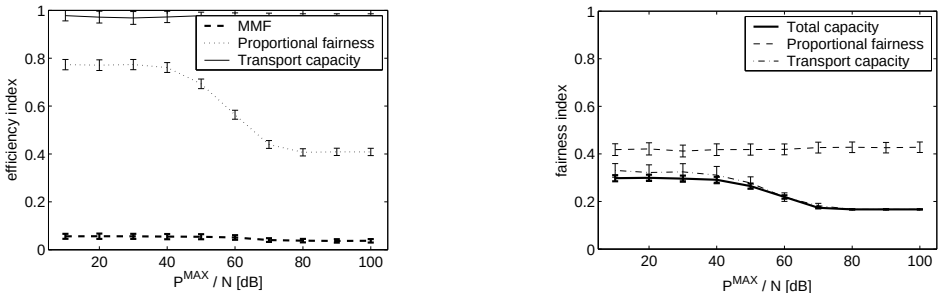


Figure 4.7: Efficiency index (top) and fairness index (bottom) of the rates that maximizes total capacity and transport capacity , and the proportionally fair ($\xi = 1$) rates. Results are obtained on random uniform network topologies.

In some existing work, like [30], maximum transport capacity was used as a design objective for a multi-hop network. An interesting question is how appropriate is this metric with respect to the rate efficiency and the rate fairness indices. In other words: Can maximizing transport capacity reconcile the rate unfairness of the total capacity objective? According to Proposition 4.4, maximizing transport capacity also exhibits high unfairness for large transmission power constraints.

We give numerical examples for realistic transmission power constraints on random uniform network topologies in Figure 4.7. We see that the rate that maximizes transport capacity is only marginally more fair and marginally less efficient than the one that maximizes total capacity . The unfairness becomes the same in both cases for high powers, as suggested by Proposition 4.4. Again, utility fairnesses represent a much better compromise between efficiency and fairness than the total capacity based metrics.

To see why transport capacity does not alleviate the fairness problem, consider a simple example with two links of distances l_1 and l_2. Suppose that a transmitter of the second link is close to the receiver of the first link, thus it is not optimal to have them both sending together, but we need to do time divisioning: the first link is scheduled α fraction of the time, and the second on $1 - \alpha$. For simplicity we assume the rate is a linear function of SNR: $R = K\,SNR$, and assume attenuation is constant, and equal to bl^{-a} where l is the distance between nodes and

$a > 2, b$ are some constants. The optimization problem now becomes

$$\max_{\alpha} \quad \alpha\, l_1\, K\, \frac{P b l_1^{-a}}{\eta} + (1-\alpha)\, l_2\, K\, \frac{P b l_2^{-a}}{\eta}.$$

Clearly, if $l_1^{-a+1} < l_2^{-a+1}$, which happens whenever $l_1 > l_2$, the optimal is to let the second link transmit all the time. We see that the unfairness properties of linear objective function cannot be compensated through weights. Although this is a very simple example, it is often found as a part of a larger network, thus we can easily find unfairness examples in arbitrary large networks. As we can see from Figure 4.7 for larger networks and longer flows, transport capacity brings only a minor improvement in fairness.

We conclude by saying that the choice of optimal ξ depends on the system design constraints and the desired trade-off between efficiency and fairness, as can be seen from the numerical results presented in this section. However, we observe that in most cases the proportional fairness ($\xi = 1$) is the most robust trade-off between efficiency and fairness.

4.8 Influence of Long-Term Average Power Constraint

In the previous sections we have seen that when we do not put constraints on battery lifetime, we have no long-term average power constraints; in the case of max-min fairness this leads to the equal rates of all flows and in the case of maximizing total capacity to zero rates of some flows.

This is not the case when the long-term average power constraint is smaller than the transmission power constraint, as it is illustrated on Fig. 4.8. When the long-term average power limit is very small, the optimal allocation is almost the same, regardless of the choice of the performance metrics. As the long-term average power limit grows, the difference becomes more significant and the unfairness of total capacity and inefficiency of max-min fairness are visible.

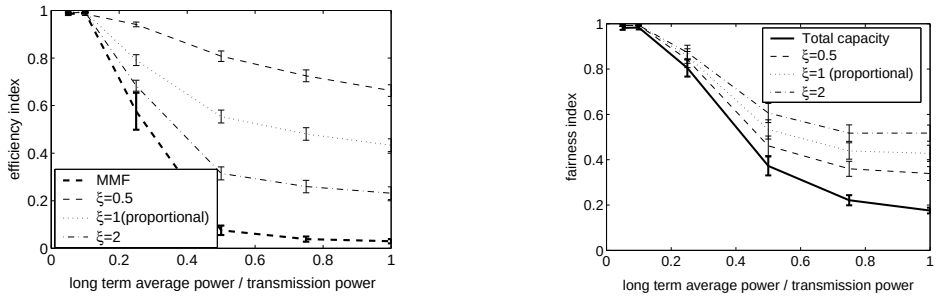


Figure 4.8: Efficiency index (left) and fairness index (right) of max-min and proportional fairness for finite long-term average power constraint.

4.9 Summary of Main Findings and Conclusions

The main findings in this chapter are:

- We proved that under a general model of an ad-hoc wireless network, in the limit of long battery lifetime, max-min fairness leads to equal long-term average rates of all flows, regardless of network topology, channel and noise variations, routing or power constraints. This means that all rates are equal to the rate of the worst flow, which makes the network very inefficient. The same happens when considering long-term average transport rates. We showed numerically that the problem persists with battery-lifetime constraints. This conclusion is in sharp contrast with the findings from the framework of wired networks, where max-min fairness is widely used. Also, this generalizes the result in [7]; it shows that their finding is not a unique property of 802.11 and that any protocol that strives for max-min fairness will have the same problem.

- We also showed that a protocol that maximizes the total capacity starves flows with bad channel conditions for sufficiently high powers. We proved analytically in this setting that if we consider a network with no random fading, only the most efficient flow gets a positive rate and all other flows have a zero rate. The same holds when maximizing transport capacity .

  We verified numerically, on a large number of random networks, that this unfairness occurs in networks with or without random fading, and not only at the limit, but also with realistic transmission power constraints. This generalizes the results in [21], showing that this unfairness property is not a problem of UWB but rather of the design criteria. We also showed that the use of transport capacity, although fairer than total capacity, does not completely alleviate unfairness and can also assign zero rates to the worst flows.

- We further showed that for very small battery-lifetimes, the max-min fair, proportionally fair and rate maximizing allocations are almost equivalent. In this case, fairness is not an issue and any of these metrics can be used in a design. However, we found that this, in general, does not hold for realistic power constraints.

- Finally, we analyzed the general form of the utility fairness, described by Equation (4.1). We showed that for values of ξ around one, the utility fairness represents a robust trade-off between fairness and efficiency, which is insensitive to different transmission power and long-term average power constraints, and network topologies. Thus the proportional

fairness (which corresponds to $\xi = 1$, and is the sum of the logarithms of the achieved rates over all source destination pairs) is an ideal candidate metric when designing or evaluating the performance of an ad-hoc wireless network.

We also presented detailed simulation results for different network topologies and different power constraints. These results can be used as a guideline for choosing a performance metric for a given design objectives. One can use the results to fine-tune the utility metric through parameter ξ in order to achieve the desired trade-off between efficiency and fairness for a given power limitations.

The findings also suggest that 802.11 should be redesigned with proportional fairness as a design objective, in order to avoid inefficiencies observed in [7]. In case of 802.11 networks with an access point, this would be similar to temporal fairness [81].

We analyzed a model of a wireless network, which includes mobility (which is implicit in the model), fading, routing, power control, scheduling and rate adaptation. Although fading, routing, and flow control may be arbitrary, we assume that scheduling, rate adaptation and power control are optimal (which corresponds to an efficient network without protocol overhead). The main reason for doing so is that we wanted to analyze the fundamental tradeoffs in the choice of a performance metric, and we wanted the analysis to be independant of the choice of the MAC protocol.

Our results suggest the kind of behaviour that can be expected of a performance metric with an arbitrary MAC. The results do not apply directly to a specific MAC and they greatly depend on the implementation of scheduling, rate and power control. However, as we show in the example on 802.11, the results are likely to hold for realistic protocols. Our findings thus give guidelines for choosing a performance metric when designing or evaluating a wireless network.

All the metrics analyzed in this chapter are rate-based performance metrics. The power constraints were considered explicitly, rather than through performance metrics. Still, powers can be incorporated in all three types of metrics analyzed here. Future work would be to analyze the ideal power-based and combined performance metric for an ad-hoc wireless network.

Chapter 5

Optimal Power Control, Rate Control, Scheduling and Routing in UWB Networks

5.1 Introduction

In this chapter we analyze cross-layer design for multi-hop ultra-wide band (UWB) best-effort ad-hoc networks. We are interested in finding a jointly-optimal design of the medium access and physical layer (MAC and PHY, namely, power control, rate control and scheduling) and routing protocols for a multi-hop, best-effort, UWB network. As explained in Chapter 4, our objective is to maximize flow rates to achieve proportional fairness, given nodes power constraints.

The specificity of UWB is expressed by the linear dependence between the rate and signal-to-noise ratio at the receiver. It is known that, in wireless networks, different routing strategies can imply differences in MAC protocol design. Hence we search for the jointly optimal routing, scheduling, rate control and power control.

We start by modeling each building block of a wireless network as a set of free variables and a set of constraints on these variables. These constraints define a set of feasible rates, that can be achieved in the network. Our goal is to find the proportionally fair rate allocation among the feasible allocations.

We solve the above optimization problem by a combination of analytical and numerical techniques, and we analyze properties of the optimal solution of the model for different networking scenarios. We find that the optimal solution is characterized by the following rules:

Exclusion Regions and Rate Adaptation - When data is being sent over a link, it is optimal to have an exclusion region around the destination, in which all nodes remain silent during

transmission, whereas nodes outside of this region can transmit in parallel, regardless of the interference they produce at the destination. Additionally, the source adapts its transmission rate according to the level of interference at the destination due to sources outside of the exclusion region.

Static Size of the Exclusion Region - The optimal size of this exclusion region can be approximated with a value that depends only on the transmission power of the source of the link, and not on the length of the link nor on positions of nodes in its vicinity.

The Optimal Power Control is $0 - P^{MAX}$ - Each node in a given time slot either sends data at the maximum power or does not send at all.

The optimal routing is MELR - As for the routing, we restrict ourselves to a subset of routes where on each successive hop we decrease the distance toward the destination, and we show that relaying along a minimum energy and loss route is always better than using longer hops or sending directly, which is not obvious since we optimize rate and not power consumption.

Insensitivity of MAC to the Choice of Routes - We fix a routing protocol and we show that different PHY and MAC protocol offer the same performance, relative to other PHYs and MACs, regardless of which specific routing protocol we choose.

For narrow-band networks, these findings do not hold, as we discuss in detail in Chapter 7. Therefore, the design of an UWB network should be addressed differently than for narrow-band.

The finding about the optimal power control is proved theoretically for an arbitrary network scenario (a detailed proof is presented in Chapter 6). The rest of the findings are demonstrated numerically. We solve the optimization problem exactly for simple networks on a line and approximately on random topologies in a plane with up to 50 nodes with various power constraints, traffic matrices, and mobility parameters.

One of the consequences of our findings is that, in the case of very-low power networks, the size of the region should be very small. In other words, in such cases it is optimal for all nodes to transmit concurrently all the time, instead of enforcing exclusions. In parallel, a source should adapt the rate to the level of interference at a receiver. We then consider a realistic example of a low-power personal area UWB network with transmitting powers of order of μWs. We find that in this case, the size of exclusion regions is approximately one meter.

We next propose a novel interference mitigation technique for the physical layer that further decreases the effect of interference. This interference mitigation diminishes the size of exclusion

regions to a few centimeters. With the modified physical layer that mitigates most of the interference, we can now apply our findings to design very simple and completely decentralize, yet optimal, MAC and PHY protocols.

In the following section we explain our motivation in detail and give the problem definition. In Section 5.3 we define the assumptions and describe our network model in detail; we translate our initial problem into a numerical optimization problem. In Section 5.5 we solve the numerical problem exactly for a static linear network. In Section 5.6 we present an approximate method for solving the general case of a mobile network in the plane and in Section 5.7 we discuss the numerical results. In Section 5.8 we show how our findings can be applied on a design of very-low power networks. Finally, in Section 5.9 we summarize our findings, and give conclusions and directions for future work.

5.2 Motivation and Formulation of The Problem

5.2.1 Physical Layer Properties

Ultra-wide band (UWB) is an emerging radio technology for wireless networks. According to the FCC, an ultra-wide band transmission has a bandwidth that is larger than 25% of the carrier frequency. There are currently several proposals of the physical layer for UWB, networks such as those based on coherent receivers [96, 2, 58] and non-coherent receivers [83].

Despite a large diversity in the existing proposals, they all have in common [93, 87] that the Shannon capacity of the optimal wide-band radio is a linear function of the SINR at the receiver. From the implementation point of view, linear rate function means that given a fixed desired error probability, the maximal achievable rate is a linear function of SINR at the receiver. This property holds for all the existing proposals (e.g. [96, 2, 83, 58]).

We also assume that the physical layer allows for **rate adaptation**: a sender uses variable coding to adapt the data rate to the channel conditions. One can use repetition coding, as in [96], or more advanced convolutional codes [32, 26, 2, 58]. By increasing the power of the code, the sender increases the effective SINR at the receiver, hence decreasing the probability of error, but at the same time decreasing the rate, and vice versa. The rate can be very fine-tuned since this is done at the symbol level.

A consequence of rate adaptation is that **an arbitrary level of interference is possible**. For a given level of interference at the receiver, a sender can tune its rate by adjusting the code, in order to achieve a desired error probability. This way, a sender can avoid collisions while still

obtaining the maximum possible rate for a given interference level. However, as we show later in the chapter, allowing for an arbitrary level of interference may still not be optimal.

Our model holds for all multi-hop wireless networks whose underlying physical layers satisfy the three conditions from above. This includes [96, 2], as explained in Section 5.3, and potentially other UWB implementations. In cases where rate can be adapted only coarsely to the signal-to-noise ratio, our solution is only approximately optimal. The UWB model of [83] differs from ours in a number of features. For more discussion see Section 5.9.

An 802.11 network does not allow for an arbitrary level of interference and adaptive coding, and we show that the MAC protocol design paradigm in our case is significantly different than the 802.11 one [30].

Although the adaptive coding techniques can be applied to a narrow-band system, the linear rate function is a particularity of wide-band systems only. This chapter is focused on wide-band systems. For more discussion on narrow-band systems (like 802.11 or CDMA), see Chapter 7.

5.2.2 Cross-Layer Design in Wireless Networks

Access to the radio medium is traditionally considered a problem of the MAC layer. However, it has recently become evident that a traditional layering approach that separates routing, scheduling, flow control, power and rate control might not be efficient for ad-hoc wireless networks [29]. This is primarily due to the interaction of links through interference, which implies that a change in power allocation or schedules on one link can induce changes in capacities of all links in the surrounding area and changes in the performance of flows that do not use the modified link. Some examples of cross-layer design are given in Chapter 2. We are interested in finding a jointly optimal routing, MAC and PHY strategies, and we formulate our question as that of a joint optimization of rate control, power control, scheduling and routing.

5.2.3 Performance Metric

Performance metrics for wireless networks can be divided in two groups: rate based and power based metrics. Here we are primarily interested in maximizing rates and we consider power control a mechanism for increasing a rate-based performance metric, rather than saving battery lifetime (though it may come as a desirable offshoot).

A detailed discussion on different rate-based performance objectives is presented in Chapter 4. There, it is shown that the most appropriate metric for a multi-hop wireless network, in which the rate function is a strictly increasing function of signal-to-noise-and-interference ratio,

is proportional fairness (defined in Section 4.2.2). The UWB physical layer has this property, and we will use proportional fairness as our design objective.

Although in this chapter we focus on rate-maximization performance metrics, we note that the power consumption is also an important performance metric for wireless ad-hoc networks. Not considering powers in a performance metric may lead to battery drain of some highly utilized nodes. However, there is a whole direction in research that studies incentives for cooperations in wireless networks. We assume here that some of these schemes are implemented in the network, hence when a node is not willing to cooperate for various reasons (such as extensive battery drain) will deny it services and will not be visible any more for nodes asking the service.

5.2.4 Routing and Mobility

The optimal route of a flow in the context defined above maximizes the utility of the system, and at the same time attains the Wardrop equilibrium [47] (the Wardrop equilibrium is the Nash equilibrium when a number of users, or in our case packets, is large). This route is not necessarily close to the shortest route as it depends on rates and routes of other flows.

In order to simplify the problem, conventional routing protocols like AODV [69] and DSR [42] focus on a set of routes that have as a next hop a node closer to the destination. We adopt this approach. We focus on routes that consist of nodes that lay on a "minimum energy and loss route" (MELR). The MELR of a flow is defined as the route that maximizes the rate of the flow in a hypothetical network with a single flow where only one node can send at a time; for a more precise definition of MELR see Section 5.4.3. We optimize routing by tuning the length of the hop. One extreme example of such a route is the MELR route itself, when the smallest possible hop lengths are used; another extreme is the direct route (DIR) where a source does not relay but sends directly to the destination.

Any implementation of routing incurs an additional penalty on network performance. There is overhead due to the implementation of the protocol. In order to maintain routes, a routing protocol needs to exchange beacons and keep-alive messages, thus consuming a part of network resources. We assume this overhead is fixed for each hop and we express it through a fraction of the rate it consumes.

Another penalty imposed by routing is due to the mobility of nodes. The longer the path is, the higher the probability that the path will break is, since one or more nodes will move. This probability depends, to a large extent, on the mobility pattern and speed, as well as on the properties of the implementation of the UWB physical layer on senders and receivers. We avoid in-depth modeling of these factors because the goal of this chapter is to anticipate to which extent

mobility influences the optimal choice of routes in an UWB network. We incorporate the impact of mobility through the probability that a packet is lost at each hop of the route.

5.2.5 Problem Formulation

We define the scheduling, routing, rate and power allocation problem in a UWB network as a numerical optimization problem, whose goal is to maximize the total utility of the system. The questions we pose are:

- **How to choose routes for flows ?** - Given an arbitrary set of nodes and traffic demands, what is the routing that achieves proportional fairness? What is the optimal hop length of a flow at each intermediate node?
- **How to organize the access to the network (MAC and PHY protocols) ?** - Given the optimal routing, when and with what rate and power should a link transmit? While a link is transmitting, how should it control its power, the interference at the receiver, and how should it adapt its rate?
- **How do routing, MAC and PHY protocols interact ?** - How do the optimal MAC and PHY protocols depend on the choice of the routing protocol and vice versa ?

Due to high complexity, it is not possible to solve this problem exactly in a general case. We solve it using a combination of analytical and numerical techniques. This is explained in more details in Section 5.4.4.

5.3 Assumptions and Modeling

5.3.1 Notations

We will use the same model and notations as defined in Section 4.4.1, with some minor modifications described throughout this section.

5.3.2 Physical Network Model

In this section we describe in details the ultra-wide band physical model of the network on which we base the design of our protocol. We first present a brief description of the Win-Scholtz physical model [96] which we use latter as an example to illustrate our modeling assumptions. However, we recall that our model is valid for large variety of UWB physical layers.

The Win-Scholtz Physical Model

We assume that the physical layer is based on the ultra-wide band radio described in [96]. This radio is based on pulse position modulation (PPM) and a coherent receiver.

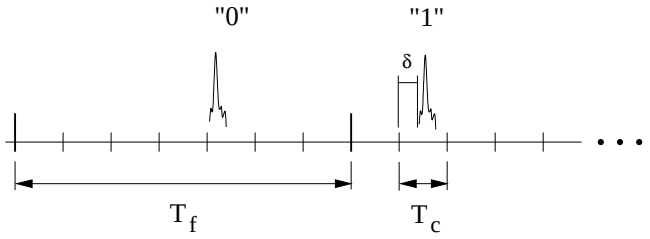


Figure 5.1: Ultra-wide band physical layer with PPM, the model of Win Scholtz [96].

Time is divided in frames of duration T_f. Each frame is divided into bins of duration T_c. A node transmits one pulse per frame, and it has a pseudo-random time hopping sequence that tells it in which bin to transmit. A time hopping code for the example on the Figure 5.1 is $\{5, 2, ...\}$. Having chosen a bin, the node sends a very short pulse within the duration of the bin. If it is sending a logical zero, this pulse is sent at the beginning of the bin, and if it is sending a logical one, the pulse is delayed by δ. On the example on Figure 5.1, the first pulse carries zero, and the second carries one.

Time hopping is used to achieve multiple access. The source and the destination of each link have a common pseudo-random time hopping sequence that is independent of other links' sequences. For other users not knowing the time hopping sequence, this signal has the statistical properties of Gaussian noise, due to randomness in time-hopping codes. It has been shown that for the particular receiver used in [96], the total noise received, comprised of background noise and a sum of signals from other active links, will be perceived by the decoder as a Gaussian noise when a number of interferers is large.

Power Attenuation

Assume the source of link l sends data in a given time slot n with transmission power p_l^n. The power of a signal at the receiver is attenuated due to wave propagation and is $p_{\text{rcv}_l^n} = p_l^n h_{ll}$, as explained in Section 4.3.1.

We also assume there is no fast random fading, since this has been shown in [87, 83] for wide-band radios. Slow random fading, which is due to mobility, is typically of a larger time-scale than a packet transmission. The mobility thus does not affect packet reception, since channel can

be estimated after each packet, but it affects a routing protocol, as explained in Section 5.3.5. Therefore, in this analysis we will assume there is the single system state ($|\mathcal{S}| = 1$).

We further assume that a transmitted signal attenuates with power law as a function of distance from the sender. If the link length is d_l then we have $h_{ll} = K_a d_l^{-\gamma}$, and the strength of the signal at the receiver $p_{\text{rcv}}{}_l^n$ will be

$$p\text{rcv}_l^n = p_l^n K_a d_l^{-\gamma}, \tag{5.1}$$

where K_a and γ are constants. This is a commonly used attenuation model for wireless transmissions, and has been confirmed as an indoor propagation model for UWB in [28]. We assume further on $K_a = 1$ since the parameter K_a influences the results only to a constant factor.

Rate Function

Let E_x be the average received power of a symbol and let η be the average power of the noise at the receiver of link l during time slot n. This noise consists of a white noise plus the total interference from other users. We approximate this noise as Gaussian, and denote the power of the interference as Z. Although it is not always true, the Gaussian approximation holds in most of the cases; it has been shown in [96] that it holds for Win-Scholtz UWB physical model in presence of a large number of interferers. The signal-to-interference-and-noise (SINR) ratio at the receiver is a ratio of received power by the total interference perceived by the receiver, including the ambient noise and the communications of other links that occur at the same time. We have $\text{SINR}_l^n = E_x/(\eta + Z)$. If there is no concurrent interfering transmissions ($Z = 0$), this ratio is also called signal-to-noise ratio (SNR).

Rate of a link can be modeled as a function $r(\text{SINR})$ of the signal-to-interference-and-noise ratio at the receiver, as described in Section 4.3.1. Information-theoretic results in [93, 87] show that a Shannon-capacity of a multi-path fading AWGN wide-band channel is a linear function of SINR

$$x_l^n = K \times \text{SINR}_l^n. \tag{5.2}$$

Thus for a given desired bit-error rate on the link, an efficient wide-band physical layer implementation should have a linear rate function within the operational interval of SINRs.

Equation (5.2) can be easily demonstrated on a widely known UWB model of Win-Scholtz [96], which is based on PPM and time-hopping. From the Eq. 8 in [22], we see that when we fix the desired bit-error rate $P(e)$ we have that $K' = N_s \times E_x/\eta$. Since $\text{SNR} = E_x/\eta$ is the signal-to-noise ratio of the received symbol, and the rate $x_l^n = 1/N_s T_f$, this yields straightforward

to Equation (5.2), with $K = 1/K'T_f$ being a function of bit-error rate and the length of time-hopping sequences.

The key difference between narrow-band and wide-band physical layers is that narrow-band ones have a strictly sub-linear (typically log) rate function. Most of our findings from this chapter hold only for a linear rate function. For more discussion on non-linear rate function, see Section 7.

The major difference between this model and the 802.11 LAN standard as modeled in [30] is that 802.11 allows only limited interference. On one hand, if the interference is high and the signal power is low, so that the SINR is below the threshold, 802.11 will not be able to receive any data. On the other hand, we assume our physical model can adapt its coding (either by using convolutional coding [32, 26] or simple repetition coding [96]) and still achieve positive rate, for arbitrary SINR value. We show later that these differences lead to a fundamental change in MAC protocol design. In the performance analysis throughout the chapter we always use our model of the physical layer, in order to allow for a fair comparison of MAC and routing protocols.

Orthogonality Factor

It has been shown in [75] that for very large bandwidth, parallel transmissions become completely orthogonal and do not interfere with each other. However, in the case of finite bandwidth system it is never the case. We introduce an orthogonality factor β that models how much of wide-band interference is captured by a receiver. The specific value of the orthogonality factor depends on the implementation of a UWB system. In the case of Win Scholtz [96] model, as shown in [22], this factor is of the order of $1/T_f$.

If η is the white noise at a receiver, and Z is the total interference from other sources, then the effective noise observed through the decoding process is $\eta + \beta Z$ and the rate of link l in slot n is

$$x_l^n = K \times \frac{p_l^n h_{ll}}{\eta + \beta Z} = \frac{K}{\beta} \times \frac{p_l^n h_{ll}}{\eta/\beta + Z}. \tag{5.3}$$

We see that the optimal architecture of such a network with white noise of intensity η is thus equivalent to the same network without orthogonality factor, but with white noise intensity η/β. Thus, without loss of generality, we can assume that $\beta = 1$.

Power Constraints

Different types power constraints are described in Section 4.3.4. In this chapter, we assume nodes have enough battery power and the only goal is to maximize rate. Hence, we will assume

there is no long-term average power constraint, and the only power constraints are imposed by hardware design and regulations.

We next illustrate how peak power constraints and transmission power constraints, defined in Section 4.3.4, can be derived for a realistic physical model, on the example of Win-Scholtz model [96]. We consider two types of limits imposed by hardware implementations and regulations: pulse power limit $P^{MAXpeak}$, and average power limit P^{MAXavg}.

Let us denote with E_x the energy and with T_c the duration of a UWB pulse of link l in slot n. Since a pulse is very short, we can assume it has a constant power E_x/T_c. This power is limited by $E_x/T_c \leq P_l^{MAXpeak}$, and $P_l^{MAXpeak}$ is indeed the peak power constraint, as it is the maximal instantaneous signal power generated by the physical layer.

In the model described in [96] a sender is allowed to send only one pulse during the duration of one frame T_f. The average transmission power during slot n is then $p_l^n = E_x/T_f$.

The power of a transmitted UWB pulse is limited by the pulse power limit as $E_x/T_c \leq P_l^{MAXpeak}$, or $p_l^n \leq P_l^{MAXpeak} T_c/T_f$. The average power limit imposes $p_l^n \leq P_l^{MAXavg}$. In the Win Scholtz model, it is assumed that the duration of a frame T_f and a chip T_c are fixed and predefined by a protocol. From the above, we have define a single transmission power constraint

$$p_l^n \leq P_l^{MAX} = \min\{P_l^{MAXpeak} \frac{T_c}{T_f}, P_l^{MAXavg}\}. \tag{5.4}$$

Note that when the transmission power constraint, defined in Equation (5.4) is satisfied, the peak power constraint is also satisfied. Therefore, we have a single power constraint in our model, which is $p_l^n \leq P_l^{MAX}$. Therefore, it fits in the model described in Section 4.4.1.

5.3.3 MAC Layer

We use the model of a MAC layer described in Section 4.3.2. Here, we define two specific examples of a schedule that we will analyze in the chapter. One is to allow only a single node to send at any point in time. We call it "total exclusion"; it can be implemented with time division multiple access (TDMA) or contention resolution protocols such as CSMA or token passing. Such a protocol is proposed in the 802.15.3a standard. Another example is a schedule in which as many nodes as possible send at the same time, as in [83].

Protocols implementing a given scheduling scheme usually involve a large complexity, and are left outside the scope of this chapter. Also, the implementation of a MAC layer typically includes ARQ related retransmission schemes. By using Equation (5.2), which describes the Shannon capacity, we implicitly assume that the MAC layer uses FEC and ARQ such that they

approach the capacity. An efficient implementation of these schemes is also out of scope of this chapter.

5.3.4 Routing

Communication is possible simply by letting each source transmit data directly to its destination. We call this direct routing (DIR). If all nodes use direct routing, we essentially have a single-hop network. The key question is whether one can improve performance by relaying data over intermediate nodes.

In a wireless setting each node can communicate to every other node, and there is an exponential number of possible relaying paths. Additionally, each source can use several paths in parallel, depending on the load on each of them, in order to achieve Wardrop equilibrium [47]. These paths depend on rates and routes of other flows, and are not necessarily close to the shortest path.

The most general model of a routing is described in Section 4.3.3. Finding the optimal routing that achieves Wardrop equilibrium is a complex task, and most of the existing multi-hop wireless networks use only single-path routes, and restrict choosing routes from a predefined subset of routes, for example a subset of routes in which each hop decreases the distance to the destination (like AODV or DSR).

In this chapter we focus on relaying over nodes that are on the minimum energy and loss route (MELR). If we consider a network with only one flow, where there can be at most one node sending at a time, then the minimum energy and loss route is the route of the flow that maximizes its rate. If the network is static and has uniform power constraint, the MELR route is equivalent to the minimum energy route (MER). For more precise definitions of MELR and MER, see Section 5.4.3.

We assume all nodes have the same routing policy. The policy is defined by the length of the hop. One example of such a route is the MELR route itself, where we use the largest number of hops and the smallest hop lengths. By increasing hop length, thus taking a smaller subset of MELR route, we can obtain intermediate paths, as shown on Figure 5.2. Finally, for very large hop lengths we obtain the DIR routing policy.

Having in mind the routing protocol described above, we can simplify the routing model, described in eq. (4.3) - (4.4). Since we consider only single-path routing, thus each flow contains only one path, and F will be an identity matrix. Eq. (4.3) - (4.4) then simplify to $R\mathbf{f} \leq \bar{\mathbf{x}}$. The matrix R is defined by the routing algorithm. Note that this definition of routing allows flows to be unicast or multicast.

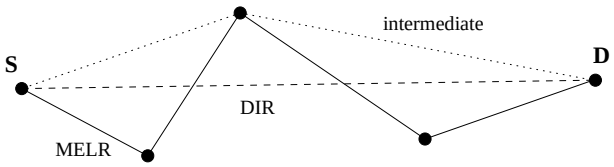


Figure 5.2: Different routing policies: DIR sends data directly from a source to a destination (dashed line); MELR relays over nodes that are on a minimum energy and loss path (solid line); intermediate routes use larger hops and relay over a subset of MELR nodes (dotted line).

Whatever routing protocol a wireless network utilizes, it will always waste some rate on the protocol overhead, such as beaconing. This in turn decreases the efficiency of a network that implements routing. We assume that each link consumes a fraction C_r of its available rate on routing, hence if the rate of a flow i is f_i^{nmr} without routing overhead, then the rate of the flow is $f_i^{\text{nm}} = (1 - C_r)\, f_i^{\text{nmr}}$ when routing is turned on (this assumes no mobility; mobility is discussed in the following subsection). In order to assess the cost of routing, we also consider a scenario without routing where a source always sends data directly to the destination. We analyze both scenarios in order to estimate what is the acceptable cost of a routing protocol in terms of rate overhead, if it is to be implemented in a network.

In the model of UWB physical layer we assume that the rate is a linear function of SINR. As mentioned in Section 5.3.2, this is true only in the operational limit of the physical layer, and can be violated if the link is too long. If a link is long and the SINR at the receiver is below the received threshold, the physical layer operates in very suboptimal region (see for example [96, 83]). If the attenuation on a link is too large, we assume that the link does not exist and we do not consider it in the routing protocol.

5.3.5 Mobility

We assume all nodes are mobile. Although we do not explicitly model nodes' movements, we describe their mobility in our model through a packet loss probability at every hop. If a node moves away, a route is broken and packets are lost. When a new node arrives, the route is reconciled, and from there on packets are transmitted. The more the nodes are mobile, the more the packets will be lost due to route destructions.

In our model we consider a snapshot of a network where nodes are not moving. This represents a "typical" topology. Let us denote with q the probability of loosing a packet at a node, or a fraction of packets that will on average be lost at each node on a long term, due to mobility.

We model the mobility through q: the higher q is, the more the nodes are mobile. Assume that the rate of flow i when nodes are not mobile is f_i^{nm}, and the number of hops on the route is r_i. Then, the rate of the same flow when nodes are mobile is $f_i = f_i^{\text{nm}} \times (1-q)^{r_i-1}$.

The mobility model considered here is on a very high-level, and might be very inaccurate if compared to more realistic models. However, a more realistic mobility model requires details of a routing protocol implementation. In this chapter we focus on the guidelines for designing the optimal protocols, and not on the actual protocol implementation details, therefore a more detailed analysis of the influence of different mobility models is out of scope of the chapter.

5.3.6 Traffic Demand and Flow Control

As in Section 4.3.3, we assume all flows have infinite amounts of data to send. Since lower protocol layers will define in a unique way the available rate for each flow, we assume our flow control layer is able to completely use this available rate.

5.3.7 Performance Objectives

As explained in Section 5.2.3, our performance objective is to find a scheduling, routing and power allocation such that the long term average flow rates are proportionally fair (for definition, see Section 4.2.2). Finding the proportionally fair rate allocation is the optimization problem that maximizes the total utility of the network and it is known to have a unique solution.

5.4 Mathematical Analysis of the Model

5.4.1 Mathematical Formulation

A general description of a set of feasible long-term average end-to-end rates is given by Eq. (4.2) - (4.2), as explained in Section 4.4.1. Applying the specificities of UWB physical layer and desired routing schemes, described in Section 5.3, we can describe set $\mathcal{F}(R)$ of feasible average flow rates under a given routing matrix R. It is the set of $\mathbf{f} \in \mathbb{R}^I$ such that there exist a schedule $\boldsymbol{\alpha}$, a set of power allocations $\mathbf{p}^n$ and a corresponding set of rate allocations $\mathbf{x}^n$ for all $n = 1 \cdots N$, and average rates $\bar{\mathbf{x}}$, such that the following set of equalities and inequalities are satisfied for all $n = 1 \cdots N, i = 1 \cdots I, l = 1 \cdots L, o = 1 \cdots O$:

$$\begin{aligned} f_i &= f_i^{\text{nmr}}(1-C_r)(1-q)^{r_i-1} \\ \mathbf{f}^{\text{nmr}} &\leq R\bar{\mathbf{x}} \\ \bar{\mathbf{x}} &= \textstyle\sum_{n=1}^{L+1} \alpha_n \mathbf{x}^n \\ \mathbf{x}_l^n &= K\,\mathbf{SINR}_l(\mathbf{p}^n) \\ \mathbf{SINR}_l(\mathbf{p}^n) &= \frac{\mathbf{p}_l^n h_{ll}}{\eta+\sum_{k\neq l}\mathbf{p}_k^n h_{kl}} \\ 1 &= \textstyle\sum_{n=1}^{L+1} \alpha_n \\ 1 &\geq \textstyle\sum_{l\,:\,l.\text{src}=o} 1_{\{p_l^n>0\}} + \sum_{l\,:\,l.\text{dst}=o} 1_{\{p_l^n>0\}} \\ \mathbf{p}_l^n &\leq P_l^{MAX} \end{aligned} \tag{5.5}$$

where $l.\text{src} = o$ and $l.\text{dst} = o$ are true if node o is the source or the destination of link l, respectively.

The goal of the problem is to find the proportionally fair rate allocation on the set of feasible rates. As follows from Section 4.2.2, it is equivalent to the following optimization problem:

$$U = \max_{R\in\mathcal{R},\mathbf{f}\in\mathcal{F}(R)} \sum_{i=1}^{I} \log(f_i) \tag{5.6}$$

where $\mathcal{R}$ is the set of possible routing algorithms defined by Section 5.4.3. The optimization problem Equation (5.6) has as free variables routing matrix R, time slots' frequencies α_n, and vectors of transmission powers assigned to links in slot n, $\mathbf{p}^n$. The values of these variables that solve Equation (5.6) define the optimal routing, scheduling and power control in the given network.

Definition 5.1 *A tuple $(R, \mathbf{p}^n, \alpha_n)$ is called optimal routing, power allocation and scheduling if it solves the optimization problem Equation (5.6).*

5.4.2 Optimal Power Allocation

Let us consider an arbitrary routing protocol, expressed through matrix R. Let us also assume that for any given power control strategy $\mathbf{p}^n$ we implement the optimal schedule α_n, such that we maximize the total log-utility of the network. Then the power control strategy which maximizes the log-utility of the network (that is, achieves proportional fairness), is that each link in every time slot is allocated either zero power or full power. This power control strategy is also called $0 - P^{MAX}$ power control. Furthermore, $0 - P^{MAX}$ is the only optimal power allocation, and any

other power control strategy is strictly suboptimal. A precise description of this statement and a theoretical proof are given in Chapter 6, in Theorem 6.2 and Theorem 6.3.

5.4.3 Minimum Energy and Loss Routes

We here define the notion of Minimum Energy and Loss Route (MELR), and give an intuitive interpretation.

Definition 5.2 *The Minimum Energy and Loss Route (MELR) of a flow is the route that minimizes*

$$\sum_{i=1}^{r} \left(P_i^{MAX} h_i\right)^{-1} (1-q)^{-r+1}$$

over all possible routes.

Proposition 5.1 *Consider a network depicted with Equation (5.5) with a single flow, and let us restrict the scheduling such that there may be, at most, one node sending at a time with maximum power. Then, the optimal routing for the flow is MELR routing.*

Proof: Consider an arbitrary route $\{l_1, \cdots, l_r\}$ from the source to the destination and let h_i be the fading on the i-th link. In the optimal rate allocation we have rates of all links equal to the rate of flow. Due to the restriction in the schedule, a rate of link i is $f = \bar{x}_i = \alpha_i K P_i^{MAX} h_i/\eta$, where α_i is the fraction of time link i is scheduled. Again, in the optimal allocation we have $\sum_{i=1}^{r} \alpha_i = 1$. From the above equation it is easy to derive that

$$f = K \left(\eta \sum_{i=1}^{r} \frac{1}{P_i^{MAX} h_i} \right)^{-1} (1-q)^{r-1}. \tag{5.7}$$

From there we see that a path that maximizes the rate in the network minimizes $\sum_{i=1}^{r} \left(P_i^{MAX} h_i\right)^{-1} (1-q)^{-r+1}$, which is indeed the MELR.

q.e.d.

In the case of a static network with uniform power constraints, the MELR route is equivalent to the well known energy route (MER), which we recall below.

Definition 5.3 *Suppose a path r consists of nodes $r(1), \cdots, r(n)$, and let us denote with $h_{i,j}$ the attenuation between nodes i and j. The minimum energy route (MER) of a flow is the one that*

minimizes

$$\sum_{i=1}^{n-1} h_{r(i),r(i+1)}^{-1}$$

over all possible routes r *for the flow.*

Furthermore, if we consider the scenario from Proposition 5.1, where the network is static and has uniform power constraints, and we fix the rate of the flow, then the MER route is indeed the one to minimize the dissipated energy.

5.4.4 Techniques for Solving the Problem

Even once we have successfully characterized the optimal power allocation, the considered optimization problem remains a highly complex optimization problem, and is difficult to solve in the general case, even using advanced optimization methods such as in [44]. A discussion for arbitrary networks with up to 6 nodes is given in [88]. It is difficult to draw general conclusions about network design from such small networks. Our method to overcome this is as follows.

First, we solve exactly a small, one-dimensional simple network. We consider a ring topology where all nodes are equally spaced and send data to their d hop away neighbor on the right. We are able to find an analytical solution for this simple model. This initial steps allowed us to identify the properties of the optimal solution that we describe in Section 5.1.

Then we next consider a random, two-dimensional network. We define a number of alternative strategies for scheduling and routing, and analyze the performance of the various combinations. The routing strategy decides how many nodes along the MELR route is spanned by one hop. The scheduling policy is one of the following four: all nodes at a time (interference is always allowed), only one node at a time (this is, in our framework, equivalent to time division multiple access, TDMA), and two strategies with exclusion around the destination. The last two strategies work as follows. Having chosen a route, each node selects a blocking distance, that is the radius of a disk around the receiving node in which there must be no other active nodes. We repeat the same procedure for each node and each flow, and then construct a schedule in a greedy manner. Then, we optimize slot frequencies in order to maximize system utility. The two strategies with exclusion differ in the way they compute the blocking radius. One optimizes the radius based on local information: next hop distance, available power and position of surrounding nodes. The other computes the radius based solely on the transmission power of the source of the link.

We numerically analyze the performance of these strategies on a set of networks with up to 50 nodes, where nodes are randomly distributed. Since UWB propagation is short range,

we expect such network sizes to be realistic. We consider uniform and non-uniform network topologies, both in the sense of node positioning, power constraints and traffic demands. We first consider a network of uniformly distributed nodes on a unit square, where half of the nodes are sources talking to a randomly chosen destinations from the other half of the nodes. We next consider heterogeneous scenarios where randomly distributed nodes all talk to a few base stations. Finally, we consider networks with non-uniform node distributions where, again, half of the nodes talk to the other half. We further consider different levels of mobility, as this implies unreliable routes and incurs additional cost to routing. We also consider networks whose nodes have non-uniform power constraints, which model different types of wireless equipment in the same network - from high-power laptops to small-power ubiquitous computing devices.

5.5 The Static Ring Case

As mentioned earlier, the above problem is difficult to solve in the general case. In order to obtain heuristics that will help us find an approximate solution for a general problem, we first restrict our attention to a static network with ring topology and with a high level of symmetry, as depicted on Figure 5.3. In order to maintain symmetry of the network, we also assume there is no mobility ($q = 0$), hence there is no routing overhead ($C_r = 0$). Arbitrary mobile networks in plane are discussed later in Section 5.6.

The ring topology can be represented as an L-sided regular polygon with distance l between nodes.

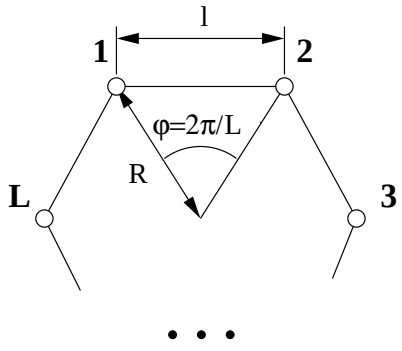


Figure 5.3: Analyzed topologies: ring.

The maximum power for all nodes is the same, and equal to P^{MAX}. Each node is a source of data, and its destination is d-hops away on the right, where $d < L/2$, thus here $I = L$.

Next we present several propositions related to the ring topologies. We prove them, and use them to simplify the numerical solution of the optimization. We then define several claims and we verify them by inspecting the numerical results on various ring sizes. These findings are then extended and verified to arbitrary networks in Section 5.6 and Section 5.7.

5.5.1 Proportional Fairness

The first proposition is about a property of the proportionally fair rate allocation.

Proposition 5.2 *In the above defined ring wireless network, if a rate allocation is proportionally fair then each flow has the same rate.*

Proof: We proceed by contradiction. Let us denote with $\mathbf{f}^*$ the proportionally fair rate allocation and assume that for some i, j, $f_i^* \neq f_j^*$. Let $(R, \alpha_n, \mathbf{p}^n)$ be the optimal routing, scheduling and power allocation that achieves $\mathbf{f}^*$.

By time division and rotation of the optimal allocation, it is possible to achieve rate allocation $\mathbf{f}$ such that

$$f_i = \frac{1}{L}\sum_{j=1}^{L} f_j^*.$$

By construction, we have that for all i, j, $f_i = f_j$, thus we have $\mathbf{f}^* \neq \mathbf{f}$.

Finally, by concavity property of the log function we further have that

$$\log(f_i) = \log\left(\frac{1}{L}\sum_{j=1}^{L} f_j^*\right) > \frac{1}{L}\sum_{j=1}^{L}\log(f_j^*),$$

which leads to contradiction.

q.e.d.

Since all flow rates are the same, we can write $f = f_i$. Due to symmetry in routing, one can easily verify that all link rates are the same, hence we can also write $\bar{x}_i = \bar{x}$. For the direct routing, a link capacity corresponds to the rate of a flow, hence the rate of flow $f = \bar{x}$. For the minimum energy routing we have d flows sharing the same link, hence $f = \bar{x}/d$.

5.5.2 Scheduling

We next describe the optimal scheduling in the ring case. In the case of MER routing, we have L one-hop links, and in the case of DIR routing we have L d-hop links. We show that the

optimal schedule consists of rotationally symmetric slots of equal lengths. For $\mathbf{p} \in \mathbb{R}^L$ and $n \in \{1, \cdots, L\}$, let us call $\mathcal{O}_n(\mathbf{p})$ a rotation of $\mathbf{p}$ such that for all $n \in \{1, \cdots, L\}$,

$$(\mathcal{O}_n(\mathbf{p}))_j = \mathbf{p}_{[(j+n-1) \bmod L]+1}.$$

Proposition 5.3 *In the above depicted ring scenario, there exists a power vector $\mathbf{p} \in \mathbb{R}^L$ such that the optimal schedule consists of L rotationally symmetric power vectors $\mathbf{p}^n = \mathcal{O}_n(\mathbf{p})$ that are equally frequent, that is $\alpha_n = 1/L$.*

Proof: Let $\bar{\mathbf{x}}$ be the vector of the optimal link rates. From Proposition 5.2 it follows that all links have the same rate, that is for all i and j, $(\bar{\mathbf{x}})_i = (\bar{\mathbf{x}})_j$.

Since $\bar{\mathbf{x}}$ is rotationally symmetric we can achieve the same rate by rotating each power allocation by an arbitrary k, hence from Equation (5.5):

$$\bar{\mathbf{x}} = \sum_{n=1}^{L+1} \alpha_n \, K \, \mathbf{SINR}(\mathcal{O}_k(\mathbf{p}^n)).$$

It follows that $\mathbf{x}^n = 1/L \sum_{k=1}^{L} K \, \mathbf{SINR}(\mathcal{O}_k(\mathbf{p}^n))$, thus all $\mathbf{x}^n$ are also rotationally symmetric. Since

$$\bar{\mathbf{x}} = \sum_{n=1}^{L+1} \alpha_n \mathbf{x}^n \leq \max_n \mathbf{x}^n,$$

we conclude that for some $\mathbf{p}$ we can represent the proportionally fair allocation as

$$\bar{\mathbf{x}} = \frac{1}{L} \sum_{n=1}^{L} K \, \mathbf{SINR}(\mathcal{O}_n(\mathbf{p})). \tag{5.8}$$

q.e.d.

5.5.3 Power Allocation

Proposition 5.4 *The optimal power allocation in a ring network is $0 - P^{MAX}$: each node in a given time either sends at the maximum power or does not send at all.*

The proposition is a consequence of Theorem 6.2 and Theorem 6.3 from Chapter 6, since a proportionally fair allocation is also Pareto efficient. For more discussion, see Chapter 6. Proposition 5.4 proves our finding 3, that all nodes should send at the maximum power when sending. Note that this result is applicable to an arbitrary network, and not only a ring network,

as stated in Proposition 5.7.

5.5.4 Routing

At this point we analyze the optimal routing policy. As already explained, we assume node i can relay over all nodes on the closest path between i and $d+i$, and the routing policy is defined by the length of the hop. If the length of the hop is 1, then data is relayed over all nodes between i and $d+i$ and this is the MER, and since the network is static, it is also the MELR route. If the length of the hop is d then the source sends directly to the destination, and the routing is called the direct routing (DIR).

Proposition 5.5 *For any network size, flow length and maximum power constraint, and under optimal scheduling and power allocation, it is the optimal to use the smallest hop length (i.e. to use the minimum energy route).*

We were not able to prove the proposition analytically. However, we solved numerically the optimization problem (5.5) for rings of various sizes and various flow lengths, and we found that MER always gave the best performance. An example of comparison between MER and DIR routes can be seen on the top of Figure 5.4 for a ring of 14 nodes, and flows of length 6.

Proposition 5.5 confirms our finding 4 for the ring case, that the optimal routing is the MELR routing.

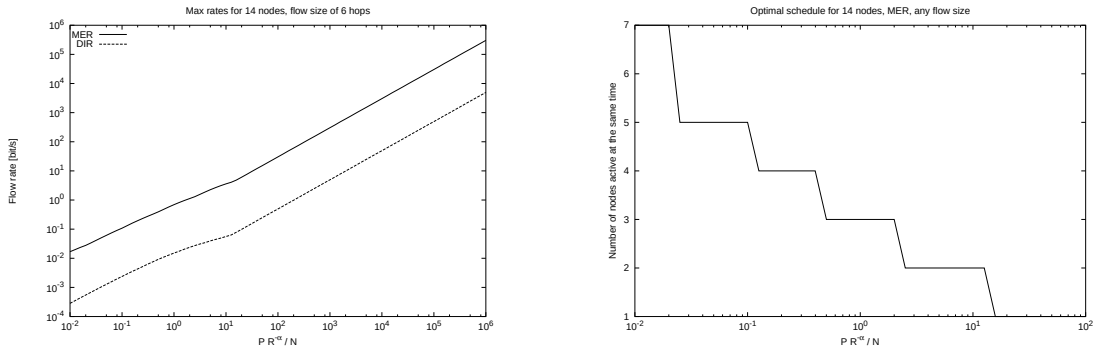


Figure 5.4: Top: maximum rates for DIR and MER routing and different values of d are depicted, for a ring of 14 nodes, versus normalized power. We see that MER routing is better then DIR for all transmitting powers constraints. Bottom: we see the optimal number of nodes active in the same slot. For low powers, 7 out of 14 nodes are active at the same time (every second node is sending). For high powers, only one node is sending at a time.

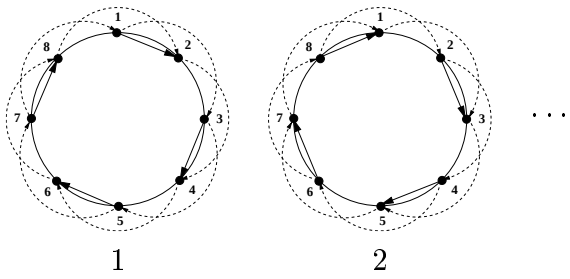


Figure 5.5: Illustration of Proposition 5.6. On the top, the optimal scheduling for MER routing is given, for $L = 8$, $d = 2$, $P^{MAX}/\eta = 10^{-2}$. Dashed arrows depict flows, and solid arrows depict active links. Numbers below are time slot numbers. Each link is either inactive, or active at the full power. There are 8 allocations in total, but the first 2 are exactly the same as the last 6. These 8 allocation repeat in a row, each taking the equal time slot, and each one is a rotation of the previous one. The distance s between active nodes is 2.

5.5.5 Exclusion Regions

Proposition 5.6 *In the optimal power allocation* $\mathbf{p}$*, each receiving node has an exclusion region around it in which all nodes are silent. The sizes of the exclusion regions of any two nodes differ by at most one. This is true regardless of the choice of the routing protocol.*

The proposition is illustrated on Figure 5.5 and Figure 5.6. Again, we were not able to prove it analytically, but we tested it for various ring sizes and flow lengths. This proposition verifies our findings 1 and 5 for the ring case. It also verifies finding 2, since we show the sizes of exclusion regions differ in at most one.

On the bottom of Figure 5.4 one can see how the number of active nodes depend on the relative transmission power $Pl^{-\alpha}/\eta$, for MER routing (which is the same as MELR in this case), where l is the distance between two adjacent nodes in the ring, and η is the intensity of the background noise. Whereas for small powers, every second node is active, for large powers only one node is active at a time. We also see that this schedule is independent of the flow length. The same holds for DIR routing.

$L = 8$, $d = 2$, $P^{MAX} = 10^{-2}$

MER	DIR
10101010	11001100

$L = 18$, $d = 2$, $P^{MAX} = 10^{-1}$

MER	DIR
100010001000100100	111000000111000000

Figure 5.6: Short representations of the optimal policies for $L = 8$ and $L = 18$: one rotation of the optimal power allocation is given, each link is denoted with 0 if inactive and 1 if active. The upper case $n = 8$ is the one depicted on Figure 5.5. In MER routing, the size of the exclusion region is smaller than the length of the hop, hence every second node is active. In DIR routing, for $L = 8$ there are 2 groups of 2 active nodes, and distance between them is 4. Again, the size of the exclusion region is smaller than the length of the hop. For $L = 18$, and MER routing, the size of the exclusion region is between 2 and 3 hop lengths l. In the case of DIR routing, some receivers have the size of exclusion region between 1 l and 2 l, and some between 2 l and 3 l.

5.6 Mobile Networks in A Plane

In this section we solve the optimization problem (5.6) for an arbitrary network on a plane. Due to the complexity of the problem we are not able to fully solve it theoretically or numerically. We are able to completely characterize the optimal power allocation and give a proof of its optimality. In order to find approximate solutions for the optimal scheduling and routing, we do not explore the full state space, but we use heuristics to find an approximate solution. We apply findings from the symmetric ring case in order to derive the heuristics. We compare these heuristics numerically on a large number of random networks in order to find the approximate solution to the problem (5.6).

We first analyze the optimal power allocation and we have the following proposition:

Proposition 5.7 *The optimal power allocation for an arbitrary network in plane is* $0 - P^{MAX}$*: each node in a given time either sends at the maximum power or does not send at all.*

As Proposition 5.4, this proposition is a consequence of Theorem 6.2 and Theorem 6.3 from Chapter 6, since a proportionally fair allocation is also Pareto efficient. For more discussion, see

Chapter 6. Proposition 5.7 proves our finding 3: all nodes should send at the maximum power when sending.

In order to find the optimal routing and scheduling, we propose a number of routing and scheduling strategies. We performed a full factorial analysis by combining these strategies, and we solved the optimization problem (5.6) for each possible combination.

First, we find the optimal route for each flow according to the chosen routing strategy. Those routes give us routing matrix R and the set of links that have to transmit in our schedule. Next we use the chosen scheduling strategy to obtain the exclusion lists, that is to say, the lists of nodes that have to remain silent while each destination is receiving.

We assign each link a weight that corresponds to the number of slots in the schedule in which it has to be active and we then use a greedy algorithm to construct a schedule. At this point, a slot represents a set of nodes that will be scheduled at the same time. In the first slot we schedule a random link, and we continue adding random links as long as they do not belong to the exclusion lists of the already scheduled links. When it is not possible to schedule further links in a given slot, we start a new slot. We repeat the algorithm until all links have been scheduled according to their weights. Note that due to the topology of a network it is possible that some links get scheduled in more slots than their weight determines (i.e. a very distant link that does not interfere with others' transmissions will be scheduled in every slot).

Once a schedule is determined, we optimize the lengths of the slots. If a link is scheduled to transmit in a given slot, according to the power hypothesis this means it sends with the maximum power. At this point we are able to calculate $\mathbf{x}^n$ link rates in each slot n from the model (5.5). Having $\mathbf{x}^n$ fixed, we optimize slot lengths $\boldsymbol{\alpha}$, which is a purely convex optimization problem, and can be solved using traditional convex programming.

We tried several heuristics to determine weights of links in constructing the schedule:

- **Equal weights:** All links have weights 1, and all get schedule approximately equal number of times.

- **Traffic weights:** Each link is assigned a weight that corresponds to the number of flows passing over it. This means a links with more flows passing over it will be scheduled more frequently.

- **Lagrangian weights:** Solving the convex optimization problem that determines $\boldsymbol{\alpha}$ for a fixed schedule, we obtain shadow prices for each component $\bar{x}_l$. We use this shadow price as the weight of a link for generating the next schedule.

We repeat the above described procedure several times for a given network. Due to the randomness in schedule generation, we obtain several different results, and we choose the maximal. In addition, in the case of Lagrangian weights, we update the weights for schedule generation at each run.

We tested different choices of scheduling weights on various random network examples. We found on all tested scenarios that changes in weights bring no difference in results, since the slot frequencies also change to compensate. This shows that the solution of our algorithm is not very sensitive to the choice of scheduling policy. Therefore, we further on used the simplest one, the equal weights strategy.

The optimization algorithm described above is implemented in Matlab, and is used to produce all the results presented in this chapter (the code can be found on [56]).

Next, we present the routing and scheduling strategies considered. Numerical results of the optimization for different networks in a plane are given in Section 5.7.

5.6.1 Routing Strategies

- **MELR routing strategy:** Each flow uses the minimum energy and loss route (MELR).
- **Intermediate routing strategy:** Flows use routes consisting of nodes that belong to the MELR path. Hop lengths in meters of all flows are the same.
- **MER routing strategy:** Each flow uses minimum energy route (MER).
- **DIR routing strategy:** There is no relaying and each flow uses the direct link (DIR).

MERL routing strategy is derived from the Proposition 5.1, where we showed it is optimal for a specific network topology. Intermediate routing strategy represents a behavior of conventional routing protocols for wireless networks and gives a large number of possible routings, including DIR and MERL routes as special cases. MER routing is equivalent to MELR on a static network with uniform power constraints. Otherwise it differs, and might not even be comprised of the same nodes as MELR.

5.6.2 Scheduling Strategies

We have seen from the ring case (Figure 5.6) that in the optimal schedule, each link should maintain an exclusion region around its destination while receiving. Other nodes outside of that region might be transmitting during the same slot. We maintain the same approach in the 2D

case, and we define two strategies that calculate the size of the optimal exclusion region s around each destination (see Figure 5.7).

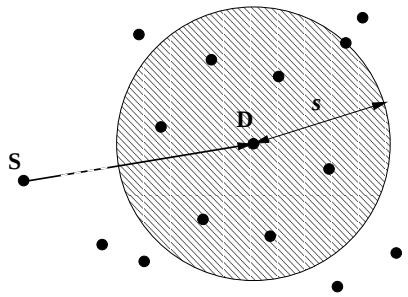


Figure 5.7: An example of the exclusion region. Node D is receiving data from node S. Its exclusion region is represented as a shaded circle, and has radius s. There are $N(s)$ nodes inside the exclusion region and they are silent during the reception. Nodes outside of the exclusion region can transmit, and the total interference they produce on D is $I(s)$.

Let us consider the destination D of link l, and assume it has an exclusion region of size s. Let us call $N(s)$ a number of nodes that are in the exclusion region of D, and let us denote with $I(s)$ the interference received by D if all nodes outside of the exclusion region would be active at the same time at the maximum power. Assume for the moment that nodes are uniformly distributed, and that all receivers have the same size of the exclusion region s. Then link l can be scheduled for transmission at approximately every $k_1 N(s)$ slot. When scheduled, it will experience the interference of approximately $k_2 I(s)$, where k_2 models the fact that not all the nodes outside of exclusion region will be scheduled at the time. The rate of link l in that case is

$$\bar{x}_l = \frac{K}{k_1 N(s)} \frac{P_l^{MAX}}{\eta + k_2 I(s)}. \tag{5.9}$$

Based on this equation, we propose the first strategy for finding the size of the exclusion region

Scheduling strategy 1: The receiver of each link maintains an exclusion region while receiving. It calculates $N(s)$ and $I(s)$ and for given k_1 and k_2 finds s that maximizes Equation (5.9). This is the size of the exclusion region. The optimal values of k_1 and k_2 are tuned on numerical examples.

Next, we propose a simpler strategy. Again, consider a network with uniformly distributed nodes of density λ, where all nodes have equal power constraints. Consider the destination D of link l. We assume it has the exclusion region of size s and we assume there is a node at the border of this region, at distance s from D. We also assume there is a power law signal attenuation and

we approximate $I(s) = P^{MAX}s^{-\gamma}$. We further approximate $N(s) = \lambda s^2\pi$, and we look for s that maximizes Equation (5.9), with $k_2 = 1$. By inspecting the first derivative, we find that the optimal s satisfies

$$\frac{P^{MAX}s^{-\gamma}}{\eta} = \frac{2}{\gamma - 2}. \tag{5.10}$$

Scheduling strategy 2: The receiver of each link maintains an exclusion region while receiving. It calculates the size of the destination region around its destination according to the following formula

$$s = \left(\frac{(\gamma - 2)P^{MAX}}{2\eta}\right)^{1/\gamma}$$

The second strategy strikes with the simplicity of a potential implementation. The left side of Equation (5.10) represents the SNR of the potential signal a destination would send to announce the exclusion region. Any node that would receive this signal with SNR higher than $2/(\gamma - 2)$ would know that it is in the exclusion region and that it should remain silent.

We also consider Total exclusion and All-at-once scheduling policies for comparison with the two strategies:

Scheduling strategy 3 (Total exclusion): Only one node may send at a time.

Scheduling strategy 4 (All-at-once): As many node as possible can send at a time, as long as each node only sends to or receives from only one node at a time.

Note that these two scheduling policies can be also viewed as policies with exclusion regions. Total exclusion is equivalent to having an exclusion region with the infinite size, while All-at-once is equivalent to having an exclusion region of size 0.

5.7 Numerical Results

In this section we present numerical solutions of the optimization problem (5.6) for various network scenarios. The goal of the analysis is to prove our findings 1-5 from Section 5.1.

5.7.1 An Example Scenario

We start by illustrating our results on a simple example, given on Figure 5.8. The network is static and power constraints are the same for all nodes. This in turn means that MER and MELR routes are equivalent. We compare the total utility (Equation (5.6)) achieved by different combinations of routing and scheduling strategies.

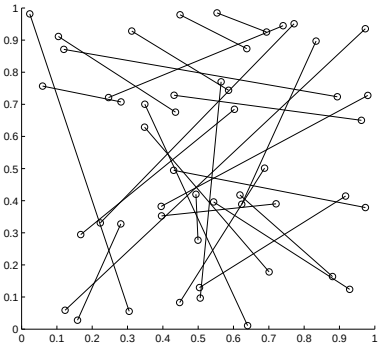


Figure 5.8: An example of a network with 50 randomly distributed nodes and 25 randomly distributed flows.

On Figure 5.9 we plot the histograms of flow rates achieved on the network from Figure 5.8 with MELR routing and different schedulings. The utility of each allocation is given in the title. We see that the utilities of scheduling strategies 1 and 2 are almost the same. We also see that the utilities achieved by Total exclusion and All-at-once schedulings in this case are significantly smaller. The worse performance in these cases can be verified from the two histograms on the right.

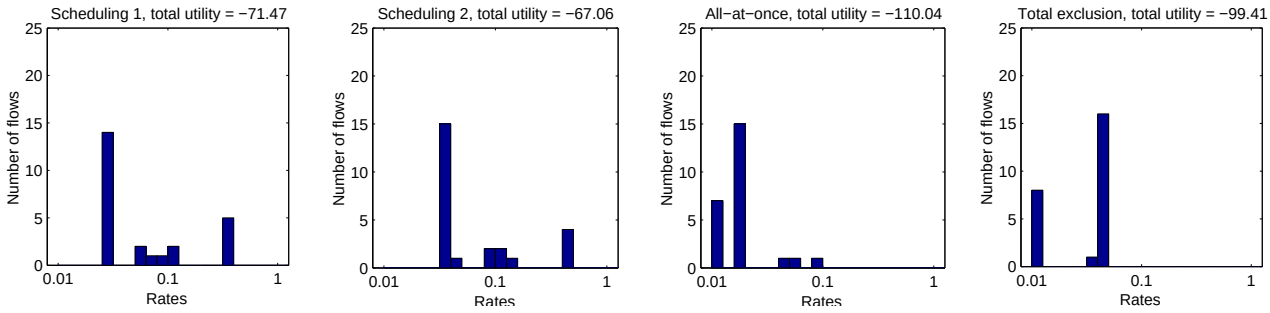


Figure 5.9: Histogram of flow rates achieved on the network from Figure 5.8 with MELR routing and different schedulings. Maximal transmission power is $P^{MAX}/\eta = -30dB$ for all nodes. The name of a strategy with the total utility achieved is given in the title of each figure. On the X axis are depicted rates in the log scales. The axis is divided into bins of equal sizes. On the Y axis are the numbers of flows whose rates belong to a given bin.

On Figure 5.10 we plot histogram of flow rates achieved on the network from Figure 5.8 for DIR and MELR routings, and scheduling strategy 2. The utility of each allocation is given in the title. We see that MELR routing is in this case significantly better than DIR.

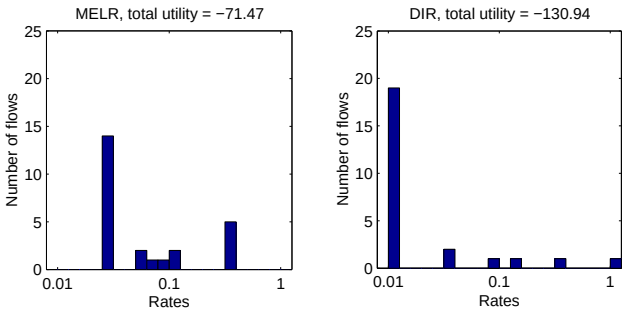


Figure 5.10: Histogram of flow rates achieved on the network from Figure 5.8 using scheduling strategy 2 and MELR or DIR routing. Maximal transmission power is $P^{MAX}/\eta = -30dB$ for all nodes. The total utility achieved is given in the title of each figure. On the X axis are depicted rates in the log scales. The axis is divided into bins of equal sizes. On the Y axis are the numbers of flows whose rates belong to a given bin.

5.7.2 Numerical Values of System Parameters

In the above examples, the actual value of P^{MAX}/η depends of the system's implementation, and on specific values of the attenuation factors, defined in Equation (5.1), and the orthogonality factor, defined in Equation (5.3). In all the numerical examples in this chapter, we have normalized the values of all the coefficients ($\beta = 1, K_a = 1$). For a realistic scenarios, one should consider $P^{MAX}K_a\beta/\eta$ instead of P^{MAX}/η (see for example Equation (5.3)).

For example, if we consider the Win-Scholtz model from [96], and take parameters from [22, 28], we obtain $\beta = 3.910^{-3}$ and $K_a = 10^{-5}$. We then have $P^{MAX}K_a\beta/\eta = -4.62dB$. Plugging it into Equation (5.10) we see that the exclusion region size is of the order of several meters. For values of these parameters in a case of a low power UWB system, and calculation of the exact size of the exclusion region, refer to Section 5.8.

Also, all the simulations are performed on a unit square. To apply the results on larger areas, one needs to scale them. For example, if we want to consider 10m x 10m square, it is equivalent considering 1m x 1m square where all transmission powers are 10^{γ} times smaller. Therefore, instead of considering $P^{MAX}/\eta = -4.62dB$ we have to take approximately $P^{MAX}/\eta = -24.62dB$. For example, if we consider Figure 5.11, for a network of 50 nodes placed on 10m x 10m square we need to take $P^{MAX}/\eta = -24.62dB$, and we see that the optimal exclusion strategy achieves significant improvements over the other strategies.

5.7.3 Homogeneous Networks with Homogeneous Traffic

We first consider homogeneous networks with 50 nodes uniformly distributed on a square of area 1. Half of the nodes are sources and half are destinations, and each source chooses randomly a destination. An example of such a network is given on Figure 5.8.

The main parameter that influences the rate is the received power over noise. Therefore, varying the density of nodes (either by changing the number of nodes or the surface of the area) is equivalent to varying the maximum transmitted power constraint. Also, an increase in the white noise is equivalent to a decrease of maximum transmission power. We consider P^{MAX}/η as the main architecture parameter of our network. We test all possible combinations of routing and scheduling strategies presented in Section 5.6, for different values of P^{MAX}/η. For a discussion on how to set realistic values of P^{MAX}/η, see Section 5.7.2.

An illustration of the comparisons is given on Figure 5.11, where we compare MELR and DIR routing combined with all 4 scheduling strategies. We find that the scheduling strategy 2 is the best one, regardless of the routing protocol and the length of hops used. Recall that the size of an exclusion zone in strategy 2 depends only on the power constraints of the corresponding source, and not on the length of the link or the positions of surrounding nodes. This confirms our finding 1 and 2 about the optimal MAC protocol. It also confirms finding 5, since neither the scheduling strategy nor the size of the exclusion region itself depend on the choice of the route.

We also compared performances of different routing protocols by varying the length of hops, for different scheduling strategies. An example can be seen on Figure 5.11 where we depict performances of DIR and MELR routings for different scheduling strategy. We found that the smallest hop length is the best, and that MELR routing is the optimal routing, regardless of the choice of the scheduling. This confirms our finding 4.

It is interesting to observe from Figure 5.11 that the scheduling strategy 1, which has more sophisticated procedure of calculating the size of exclusion regions, has a performance equally as good as strategy 2. We also see that Total exclusion is equally as good as the scheduling strategy 2 for high power constraints, and All-at-once is as good for low power constraints. This is in accordance with the construction of the exclusion region of strategy 2. On one hand, as the power is growing, the exclusion regions are growing as well, until each exclusion region occupies the whole network, and the policy becomes Total exclusion. On the other hand, when power is sufficiently small, exclusion regions will be small enough so that they do not include any nodes, and the policy becomes All-at-once.

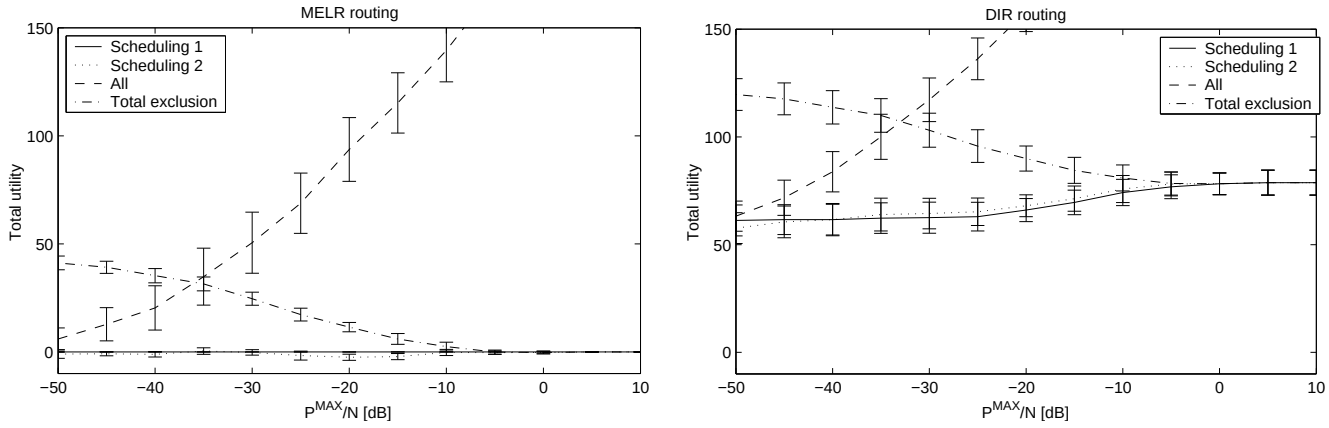


Figure 5.11: Utilities of different routing and scheduling algorithms applied on homogeneous random networks and different maximal power constraints. On the left different schedules are applied with MELR routing and on the right with DIR routing. On the x axis a relative power constraint P^{MAX}/η is given. On the y axis the differences between total utility of the reference approach and the analyzed approach are given. Here, scheduling strategy 2 with MELR is taken as the reference approach.

5.7.4 Non-Homogeneous Networks

During the derivations of our strategies we always assumed networks were rather homogeneous: nodes are uniformly distributed, routing matrix is homogeneous, and power constraints are the same for all nodes. In this section we analyze if our conclusions are also valid in non-homogeneous networks. We address three types of inhomogeneity: non-homogeneous nodes distribution, non-homogeneous traffic matrix and non-uniform power constraints.

We first consider non-homogeneous node distributions. We now assume that 50 nodes are distributed on the unit square such that 40 nodes are placed uniformly on the left half of the square and the remaining 10 nodes are placed on the right half. Sources and destinations are still chosen uniformly among nodes.

Next, we consider non-uniform traffic matrix. We uniformly place 50 nodes on a unit square, and all of them talk to the nearest base station. We analyzes cases with 1 (placed in the center of the square) and 4 base stations (placed in the centers of four quarters of the square).

Finally, we consider networks with uniformly distributed nodes and source destination pairs, but non-uniform power constraints. Each node has the power constraint randomly chosen from the interval $[(1-c)P^{MAX}, (1+c)P^{MAX}]$. This way the average power constraint of all nodes is still P^{MAX}.

In all those cases, we analyze all combinations of scheduling and routing strategies, and we

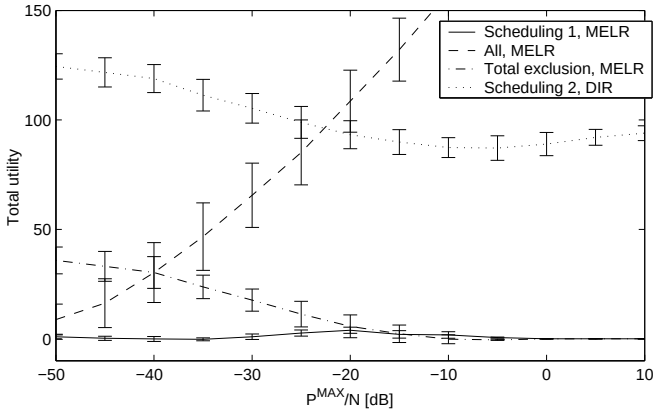


Figure 5.12: Utilities of different routing and scheduling algorithms applied on random networks with non-uniform node distribution. We assume 40 nodes are uniformly distributed on the left half on the unit square, and 10 nodes on the right half. Source-destination pairs are uniformly chosen. On the x axis a relative power constraint P^{MAX}/η is given. On the y axis the differences between total utility of the reference approach (scheduling strategy 2 and the MELR routing) and the analyzed approach are given.

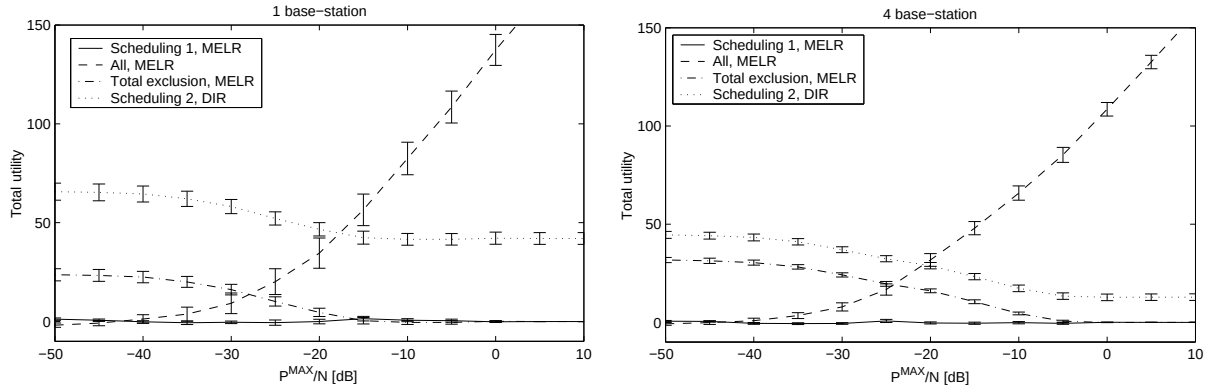


Figure 5.13: Utilities of different routing and scheduling algorithms applied on homogeneous random networks with base-stations. We assume 50 randomly distributed nodes are sending data to the nearest of 1 (on the left) or 4 (on the right) base stations. On the x axis a relative power constraint P^{MAX}/η is given. On the y axis the difference between total utility of the reference approach (scheduling strategy 2 and the MELR routing) and the analyzed approach are given.

search for the one that has the highest utility. We repeat the procedure for several random network topologies and for different values of power constraints. Examples of comparisons can be seen on Figure 5.12 for non-homogeneous network topologies, Figure 5.13 for non-homogeneous traffic matrix, and Figure 5.14 for non-uniform power constraints.

For all considered network scenarios we derive the same conclusions. We find that the optimal routing is the one with minimal hop length, which is MELR routing, regardless of the choice

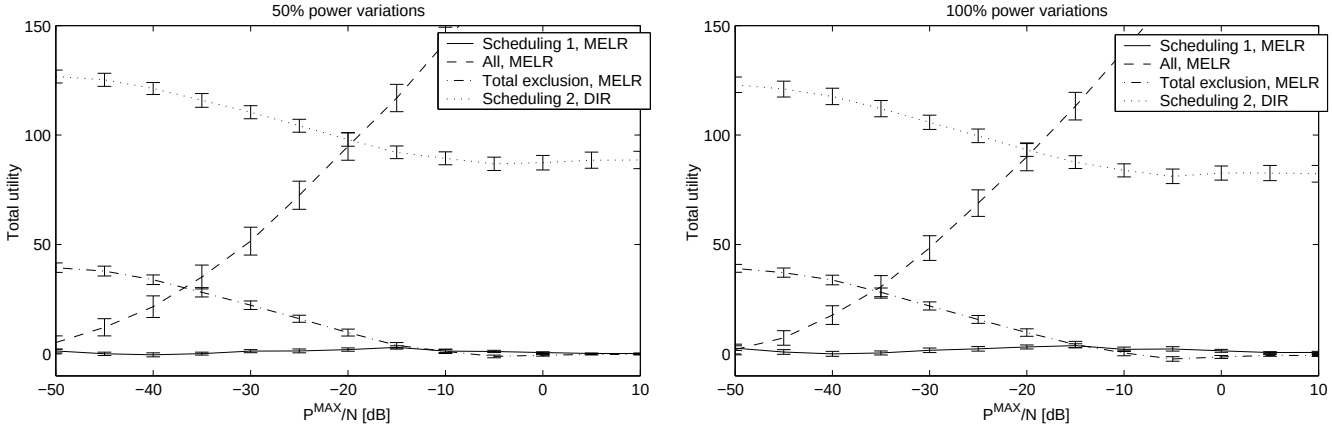


Figure 5.14: Utilities of different routing and scheduling algorithms applied on homogeneous random networks with non-uniform maximal power constraints. We assume power constraints are uniformly distributed in the intervals of $[0.5P^{MAX}, 1.5P^{MAX}]$ (on the left) and $[0, 2P^{MAX}]$ (on the right). On the x axis a relative average power constraint P^{MAX}/η is given. On the y axis the differences between total utility of the reference approach and the analyzed approach are given. Here, scheduling strategy 2 and MELR routing are taken as the reference approach.

of the scheduling protocol. This confirms finding 4. We also find that the optimal scheduling strategy is strategy 2, regardless of the choice of the routing protocol. This confirms findings 1 and 2. Again, from the property of strategy 2 that the exclusion region depends only on the power constraint it follows that the optimal MAC protocol is independent of the routing strategy, which confirms finding 5. We conclude that although our findings were initially derived for homogeneous networks, they are robust to changes in network characteristic and remain valid in non-homogeneous network scenarios.

Other interesting remarks can be made for the non-homogeneous traffic case, depicted on Figure 5.13. As we concluded, DIR routing is always worse than MELR, but it becomes more efficient as the number of base-stations increases, which is due to the fact that routes are are becoming shorter. Also, in the case of a network with non-uniform power constraints, depicted on Figure 5.14, we find that, though MELR and MER routings do not give the same routes when power constraints are not uniform, in our tests both routings gave almost equal performance.

5.7.5 The Effects of Mobility and The Cost of Routing

As discussed in the introduction, routing in the wireless networks might be expensive due to its cost, and to the mobility of nodes and instabilities of routes. In this section we investigate the

impact of mobility on our findings. Again, we test all possible combinations of scheduling and routing strategies and find which one has the highest utility. We do the test on homogeneous and non-homogeneous network scenarios described above. We considered two levels of mobility, with packet loss probabilities of 10% and 25%.

We obtain the same results in all analyzed cases. The routing with minimum hop length is the optimal one, regardless of the choice of scheduling. Scheduling strategy 2 is the optimal one for all routing strategies. As already discussed above, this confirms our findings.

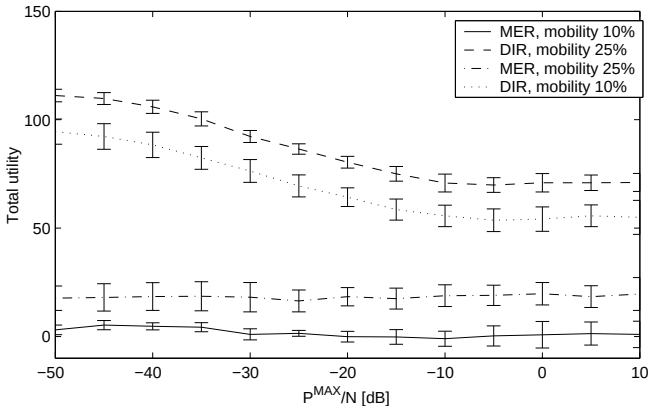


Figure 5.15: Utilities of different routing algorithms applied on mobile homogeneous random networks and different maximal power constraints. Fraction of packets lost on each hop is 10% or 25%. On the x axis a relative power constraint P^{MAX}/η is given. On the y axis are the differences between total utility of the reference approach (MELR routing and scheduling strategy 2) and MER and DIR routing again with the optimal scheduling strategy 2.

A snapshot of numerical results can be seen on Figure 5.15, where we show utilities of different routings with the optimal scheduling, for different mobility levels. As noted above, MELR routing is the optimal one, and for a static case, it is equivalent to MER routing. As the mobility level increases, MER routing deteriorates, and DIR improves. We verified that for very high packet loss ratios due to mobility DIR will eventually become equivalent to MELR. However, this does not happen in realistic examples. Furthermore, as expected, mobility does not influence the optimal scheduling strategy as it does not change behavior in any other way but adding additional cost to the routes.

Another previously mentioned drawback of routing is its cost. We next discuss what the maximal cost of a routing is after which our findings do not hold any more. We say that a routing protocol consumes a constant fraction C_r of rate of each link. It is obvious that it is not going to impact the optimal scheduling strategy, hence findings 1, 2 and 5 will continue to hold. On the contrary, a sufficiently high value of this additional cost can penalize routing such that no routing

(that is to say DIR) becomes the optimal strategy, thus invalidating finding 4. We are interested in finding for which values of C_r this will happen.

According to Equation (5.5) we have that the rate of a flow i is $f_i = f_i^{\text{nr}}(1 - C_r)$, where $\mathbf{f}^{\text{nr}}$ are the rates in the same network without routing costs. If a network consists of I flows, then the utility of the network with routing will be $U = I\log(1 - C_r) + U^{\text{nr}}$, or

$$C_r = 1 - \exp((U - U^{\text{nr}})/I). \tag{5.11}$$

For example, for a network with 25 flows, if the routing cost is $C_r = 0.25$ then the decrease in utility comparing to the same network without routing cost is going to be around 7. By inspecting the above presented graphs, and using Equation (5.11), we estimate for which values of C_r our finding 4 will cease to hold.

We first consider static networks. From Figure 5.11 we see that the difference in utilities between DIR and MER (which is here equivalent to MELR) is between 70 and 90. From Equation (5.11) we see that DIR becomes better when the cost of routing is $C_r \approx 0.85$. As 85% of the overhead can be considered very high for a routing protocol, we conclude that in realistic scenarios MELR is the optimal routing, and our finding 4 still holds. The same result holds for non-homogeneous networks.

Next, we consider mobile networks from Figure 5.15, with a packet loss probability of 25%. The difference in utilities between DIR and the optimal routing is more than 25. Again from Equation (5.11) we conclude that DIR will become optimal for cost $C_r > 0.63$. Both a packet loss ratio of 25% and a routing overhead of 63% still represent extreme values and we again conclude that, in realistic scenarios with mobility, our finding 4 still holds.

5.8 Application to Very Low-Power UWB Networks

In this section, we apply some of our findings to design a MAC protocol for a very low-power UWB network. As suggested by our findings, the size of exclusion regions depend only on transmission powers. When transmission powers are low, the exclusion region is small. Once the exclusion region becomes very small, it is not necessary any more to impose exclusions since there are no other nodes in exclusion regions. Instead, the optimal MAC and PHY should enforce all nodes to transmit concurrently and all sources to constantly adapt rates to the level of the interference at receivers. The goal of this section is to inspect if this conclusion can be applied on existing proposals of very-low power UWB networks.

We consider a network of very low-power devices, inspired by 802.15.4a standard proposal [1], whose average transmitted power is of the order of 1μW. A network is based on a pulse-position modulation (PPM) UWB layer, which is described in [58]. This physical layer is an improved version of the physical layer described in Section 5.3.2 and it has a different rate function. However, the model presented in this chapter can still be used as an approximation of the network behavior.

In this example, we focus on personal-area networks where devices talking to each other are in direct ranges, and we assume all sources transmit directly to their destinations and there is no relaying (nodes always use DIR routing). We also assume the networks are decentralized. There is no master node that defines a schedule, but different nodes contend independently for access to the channel. The goal is to optimize physical and MAC layer design for this kind of network.

Our finding about the insensitivity of MAC to the choice of routes suggests that the optimal MAC and physical layer design does not depend on the choice of the routing. Hence our findings about the optimal MAC and physical layer hold even when we apriori fix routing to DIR.

Next, the finding about power control suggests that whenever a node transmits, it should transmit with the full power, or else it should remain silent. Therefore, there is no need for power control.

Finally, each destination should maintain an exclusion region around itself whenever it is receiving data. While nodes inside the exclusion should remain silent, nodes outside of it should transmit concurrently. At the same time, a source and a destination should adapt the transmission rate to the level of interference generated by these nodes that are outside of the exclusion region.

The size of the exclusion region depends only on the transmission power, and not on the size of a link or on positions of other nodes. In our case the transmission power is 1 μW. Using Equation (5.10), and typical UWB channel statistics, given in [28], we obtain that the optimal size of an exclusion region is approximately 1-2m.

In order to verify the above result based on an approximative model of the physical layer, we used a simulator, described in [58], which has a detailed, symbol-level implementation of the physical layer. We consider a cylindric scenario from Figure 5.16. We consider cylinders with 4 and 32 nodes, and fix $l = 24$m. We vary the distance d to adjacent interfering nodes. We consider two transmission scenarios. The first is without mutual exclusion, where all senders send at the same time. The second scenario is with mutual exclusion, where only sources from one edge of the cylinder transmit concurrently. We verified numerically that this is the optimal mutual exclusion schedule for cylindric example.

As can be seen from Figure 5.17, when $d \leq 1$m, having all nodes send at the same time is

not optimal. This means that the radius of the exclusion region is indeed approximately 1 m. If an interfering node is closer than 1 m to a receiving node, the interference will significantly deteriorate the performance of the receiver and an exclusion is necessary.

It is a realistic scenario to have one or more interfering nodes at 1 m from the receiver. Still, an implementation of exclusion regions is tedious, requires a complex MAC protocol and a large overhead, which is clearly undesirable in a low-power network. Instead, we propose an interference mitigation technique, presented in the following section, that will alter the physical layer and further decrease the size of the exclusion region. This will render mutual exclusion no longer necessary even for $d \leq 1$m, and will greatly simplify the MAC layer.

5.8.1 Interference Mitigation

In our context, the cause of interference is twofold. There is the thermal noise and there are collisions between a pulse from a source with pulses from one or several interferers. The probability of collision depends on the size of a frame T_f and a size of a chip T_c (defined in Section 5.1). A large T_f/T_c ratio implies a low probability of collision. For a transmission power of 1μW, we have a typical value of $T_f/T_c = 280$. If there is only one interferer, the probability of a symbol level collision is below 1%.

It might seem that with such a low probability of collision, the effect of the interference is very small and can be neglected. However, if an interferer is very close, within the radius of the exclusion region, the energy of an interfering pulse will be much larger than the energy of the received signal. A typical decoder decodes a large sequences of received symbols together. A colliding pulse will thus impact the decision on a large number of bits, with the probability of error being larger as the energy of the colliding pulse increases. Therefore, even with a low probability of collision of less than 1%, the impact of interference is huge when the interferer is in the exclusion region, as predicted by our findings, and verified by simulations on Figure 5.17.

In order to combat the interference coming from the nodes within the exclusion region, we exploit the fact that the received power of a pulse from an interferer within an exclusion region is much larger than the power of a pulse received from the source, in order to greatly reduce the effect of interference. This *interference mitigation* mechanism is inspired by the work in [83], and represents a very simple form of multi-user detection.

Assume a source S communicates with a destination D in the presence of a nearby interferer X. If a pulse from X collides with a pulse from S, the received energy is much higher than the intended received power from S. We use a *threshold* demodulator at D that detects when the received energy is larger than some threshold $\mathcal{B}$. In this case, we skip the chip and declare

an erasure. With our choice of $T_f/T_c = 280$, the probability that an erasure occurs is very low. The loss incurred by those erasures can mostly be recovered by our channel codes, unlike losses from collisions, which are difficult to compensate. The resulting rate reduction with interference mitigation is much smaller than if we do not use it and let the channel code attempt to recover the errors created by a pulse collision. A small erasure probability translates into a small reduction of the rate. We simulate the physical layer in Matlab and evaluate the performance of our interference mitigation in the scenario of Figure 5.16. The results are depicted in Figure 5.17.

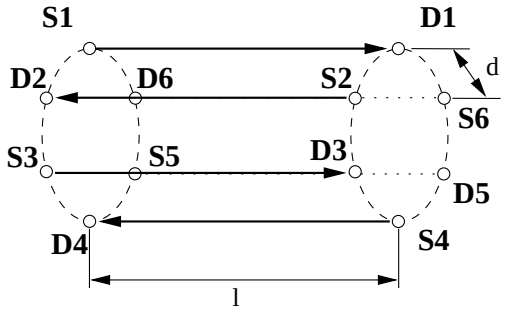


Figure 5.16: Multiple interferers scenario: n nodes are symmetrically distributed on the edges of a cylinder. Every second link is inverted such that each destination is close to an interfering source. l is distance between a source and a destination, d is the distance between a destination and the adjacent interfering source. Number of links is $n/2$ (on the figure $n = 6$)

With the interference mitigation, we change the performance of our physical layer. The effect of the mitigated interference is highly non-linear, and the effect of an interfering pulse will depend to a large extent on the distance of that interferer to a destination. Specifically, if the interferer is sufficiently close to the destination, the interference will be mitigated. Otherwise, it is not possible. Therefore, a rate of a link can not be expressed as a function of SINR only, but will depend on locations and transmit powers of each of the interferers, and the physical layer model from Section 5.1 is not applicable. Consequently, the findings from this chapter cannot be directly applied.

In order to verify our findings and to estimate the size of the exclusion region, we turn again to the cylindric example from Figure 5.16. We see from Figure 5.17 that for the case with interference mitigation, the achieved rates never fall below the rates obtained by mutual exclusion. Thus, interference mitigation renders the size of the exclusion region negligible and significantly increases the achieved throughput in the case of a close interferer.

Note that the optimal value of the threshold $\mathcal{B}$ depends on both the power of the interferer and the white noise. A too large $\mathcal{B}$ is equivalent to the case without erasures, whereas a small $\mathcal{B}$ will declare too many erasures. Our goal is to set $\mathcal{B}$ such that the erasures are declared only

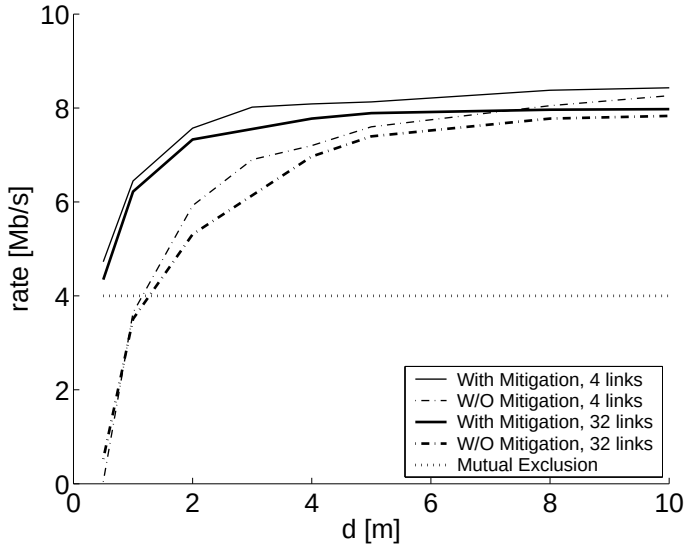


Figure 5.17: Rate achieved in the multiple interference scenario, for our physical layer (with and without interference mitigation) and mutual exclusion versus distance d. Link lengths $l = 24$, $n =$4 and 32.

due to collisions, and not due to the white noise. We found by simulation that a suitable value is $\mathcal{B} = 3\mathcal{N} + \mathcal{A}$ where $\mathcal{N}$ is the average white noise power and $\mathcal{A}$ is the estimated received signal power. Using erasures as a way to mitigate interference is suitable when T_f/T_c is rather large. However we found that our method continues to perform well for smaller values of T_f/T_c (down to 100), i.e. for higher power, but a detailed investigation is for further research, as well as the determination of the optimal value for the threshold $\mathcal{B}$.

We have seen in this section that the optimal MAC protocol for the low power UWB system, described in Section 5.3.2, needs to enforce mutual exclusion for sources and destinations that are less than 1 m away. An implementation of mutual exclusion requires a complex protocol, which becomes even more difficult in the case of UWB since it is not possible to perform carrier sensing. We modify the physical layer instead; we proposed an interference mitigation technique that decreases the destructive impact of the interference. Similarly to exclusion regions, the proposed interference mitigation technique is a way to manage interference. It is efficient in managing the interference coming from nearby nodes, and it obsoletes exclusions. We verify that with the modified physical layer the size of the exclusion region is reduced to a negligible value, which in turn greatly simplifies our protocol design.

A protocol implemented along these lines is called DCC-MAC. Its implementation is out of scope of this thesis. For more details, see [58].

5.9 Summary of Main Findings and Conclusions

In this chapter, we have answered questions on what objective the MAC layer and the routing protocol should have in a multi-hop UWB ad-hoc network. We have presented a general model for the joint scheduling, power allocation and routing optimization problem. Based on the solutions for this optimization problem, we have identified findings that we showed to hold quite generally in arbitrary UWB networks. Our main findings can be summarized in the following:

Exclusion Regions and Rate Adaptation - It is optimal to have an exclusion region around the destination of an active link. All nodes in the exclusion region have to remain silent. However, nodes outside of the exclusion region can transmit data in parallel. The destination will thus experience interference only from nodes outside of the exclusion region, and will adapt the rate accordingly. This is very much in contrast to the 802.11 CSMA/CA mechanism, where the size of the exclusion region is adapted such that the interference at the receiver is smaller than a threshold while the rate is kept fixed.

Static Size of the Exclusion Region - In most of the cases, the optimal size of this exclusion region around a destination can be approximated with a value that depends only on the transmission power constraint of the corresponding source, and not on the link length nor on the density of nodes around the destination. Although there are some counter-examples where this finding does not hold, it is approximately true for most network topologies.

Sending at Full Power - When data is sent over a link, the sender should transmit with the maximum allowed power. Otherwise, the link should remain silent. We also call this power control strategy $0 - P^{MAX}$power control. This is again in contrast with 802.11, where a source can use only one of a few modulations to send data [46]. The finding is generalized and elaborated with more details in Chapter 6.

Minimum Energy and Loss Routing - Within the routes we consider, the routing that maximizes total rate utility in a UWB network is the minimum energy and loss route itself. In addition, if maximum power constraints are uniform, and nodes are not mobile, this optimal route coincides with what is usually called the minimum energy route (MER), described in Section 5.4.3. This result is similar to the findings in 802.11 networks, as shown in [46].

Insensitivity of MAC to the Choice of Routes - Another significant conclusion is that the optimal MAC protocol does not depend on the choice of routes. We found that if all nodes

apply similar routing strategies, then the MAC protocol should maintain exclusion regions of the fixed sizes, regardless of the choice of hop lengths made by routing. This leads to an important property: a UWB network can be organized in a traditional layered manner, where a routing protocol chooses routes, and a MAC protocol organizes medium access regardless of the choice of routes. This conclusion is in contradiction with the findings for 802.11 [89] and narrow-band networks in general (see Chapter 7), and it facilitates the design of a UWB network protocol.

Optimal MAC and PHY Design for Very-Low Power Networks - For small power constraints, hence low-rate networks, we can schedule as many nodes as possible to send at a time, even if it increases interference. This finding suggests that a simple MAC protocol should be used in this type of network, as only the sender and the receiver of each link need to coordinate separately. We confirmed the findings on a realistic setting, described in [58]. We have first introduced an interference mitigation technique to the PHY, based on an extremely simple multi-user decoding approach. With such a modified PHY, we find that simple MAC and PHY protocols, that do not enforce exclusions but only adapt rates, are indeed optimal.

Our finding on the optimal power allocation is proved analytically for an arbitrary network topology or routing strategy. The detailed proof of the finding is given in Chapter 6. The rest of the findings are demonstrated numerically. We have solved analytically the scheduling and routing optimization problem for symmetric one-dimensional networks, and solved approximately by numerical simulations for arbitrary networks in a plane.

The findings give directions for implementations of routing and a MAC protocol. A routing protocol should be based on a distributed shortest path algorithm considering the inverse links' attenuations as the costs of links. To this end, any standard ad-hoc network routing protocol (AODV, DSR,...) could be used to calculate the shortest path to the destination. The cost of a link can be measured and updated during the transmission of every packet.

We also find that the optimal MAC protocol in an UWB network should be a combination of rate adaptation and mutal exclusion. The size of the exclusion region should be adapted to the parameters of a network. The fact that the optimal size of the exclusion region is defined solely on local information facilitates MAC protocol design. In that respect, our findings suggest that there are fundamental reasons why re-using MAC protocols originally designed for narrow-band (as envisioned by IEEE 802.15.3a) might not be a good idea. Details of the actual implementation of such a protocol is for future work.

Another application lies in the area of low-power UWB networks (as proposed by 802.15.4a). Our results suggest that, for sufficiently low powers, the exclusion regions are going to be small enough that no exclusion protocols are necessary. We showed on a realistic setting [58], that a simple all-at-once scheduling strategy, without complex exclusion-region based signaling, is optimal for a class of low-power UWB networks.

Application of Findings to Different Physical Layers

As we mentioned in the introduction, there exist several UWB physical models, and we have primarily focused on the Win Scholtz model [96]. For our analysis, the most important property of this physical layer is that a rate can be expressed as a linear function of SINR.

There are other physical layers where this assumption does not hold. For example, the modified WinScholtz physical layer, presented in Section 5.8, which is augmented with interference mitigation, does not have this property. Due to the non-linearity of interference mitigation (its effect on an interferer depends on the distance to that interferer), we cannot express rate as a function of SINR. However, as shown in Section 5.8, our findings approximately hold even in this case.

It would be interesting to discuss if our findings could be extended and applied also to another emerging UWB model, the model of Souilmi, Knopp and Caire [83]. Their model differs in its coding scheme and transceiver architecture. A specificity is that it implements exclusion at the physical layer, thus physical layer signaling and scheduling cannot be separated as they are in our model. Hence our findings do not apply verbatim to their model.

There are indications however that findings similar to ours might hold in [83]. Indeed, in [83], the authors have shown that the exclusion mechanism has to adapt to the number of nodes in the surrounding, but they have not analyzed it in further detail. Also, one may think that, since the behavior of an active link does not have to change due to a transmitting node that is very far away, the exclusion mechanism in [83] would in fact be needed only for nodes that are not too far apart. This resembles our findings about the exclusion regions and their optimal sizes. It remains for future work to refine our model in order analyze this generalization in detail.

Furthermore, the findings do not hold for narrow-band physical layers like CDMA or 802.11, as we show Chapter 7. Thus the fact that the rate is a linear function of the signal-to-noise ratio is important. However, although our findings do not directly hold in narrow-band network, a similar analysis of a narrow-band network could be performed to improve the existing CSMA/CA protocols for narrow-band networks (i.e. 802.11). This is discussed in detail in Chapter 7.

Chapter 6

Power Control is Not Required for Wireless Networks in the Linear Regime

6.1 Introduction

In the previous chapter we have presented several findings about the optimal design of UWB MAC and PHY layers. One of the findings was on power control. Here, we consider the problem of the optimal power control in a more general setting.

We consider again the design of optimal strategies for joint power adaptation, rate adaptation and scheduling in a multi-hop wireless network. Most existing strategies control either power and scheduling, or rates and scheduling, but not all three together as we do. We assume the underlying physical layer is in the linear regime (the rate of a link can be approximated by a linear function of the signal-to-interference-and-noise ratio), like in time hopping UWB (TH-UWB) and low gain CDMA systems, and that it allows fine-grained rate adaptation, like in 802.11a/g, HDR/CDMA, TH-UWB. The goal is to find properties of the power control in an optimal joint design.

Our main finding is that optimal power control is simple $0 - P^{MAX}$power control, i.e. when a node is sending it uses the maximum transmitting power allowed. We consider both high rate networks where the goal is to maximize rates under power constraints and low power networks where the goal is to minimize average consumed power while meeting minimum rate constraints. We prove analytically that in both scenarios the optimal allocation can always be attained with $0 - P^{MAX}$power allocation. Moreover, we prove that, when maximizing rates, and if power constraints are on peak and not average, $0 - P^{MAX}$is the only optimal power control strategy, and any other is strictly suboptimal.

In the next section we explain our motivation for this work. In Section 6.3 we describe system assumption. In Section 6.4 we give a mathematical formulation of the model of a network. In Section 6.5 we present our main findings. In the last section we summarize the main findings and give conclusions and directions for further work.

6.2 Motivation

6.2.1 Power Control and Optimal Wireless MAC Design

The first wireless MAC protocols for multi-hop networks were designed to control only medium-access. A typical example is the original 802.11 MAC. It always uses maximum power for transmitting a packet and aims to establish communication on a fixed, predefined link rate. Then several improvements to the initial approach were proposed. According to the type of improvement, the MAC protocols can be divided globally in two groups. The former group of protocols [81, 35, 58] is focused on rate adaptation: the transmission power is still kept fixed, but the rate is adapted to the actual channel conditions and the amount of interference. The latter group of protocols [45, 60, 64, 61] considers power adaptation while keeping the rates fixed. However, there are no MAC protocols that adapt both rate and power at the same time, and the fundamental issues in this joint adaptation problem are not well understood. In this chapter we make a first step by showing that, perhaps contrary to intuition, there is a whole class of networks (those operating in the linear rate function regime, see paragraph 6.2.3) for which power control is not required, or may even be suboptimal.

We consider a wireless network with arbitrary scheduling, rate adaptation and routing strategies, and we are interested in characterizing the properties of the optimal power allocation strategy in this setting.

6.2.2 Rate Adaptation and Rate Function

The physical layer of a wireless link defines communication parameters such as bandwidth, modulation and coding, which can be used to establish communication with some level of bit or packet errors. One of the most important parameters of the physical layer is signal-to-interference-and-noise ratio (SINR) at the receiver. The higher the SINR is, the higher communication rates can be attained, and one of the goals of networking design is to efficiently track and adapt SINRs and/or rates on links. The function that gives the maximum achievable rate for a given SINR is called the *rate function*. We discuss this issue in detail in Chapter 4 and Chapter 5.

6.2.3 Linear Regime

The rate function of an efficiently designed system is a concave function of SINR. Furthermore, in many cases, especially when bandwidth is large or the target SINR is low, it is a linear function. Some examples of physical layers, where rate function is linear, are TH-UWB [96] and low or moderate-gain CDMA [19]. These physical layers are in linear regime in the whole operational SINR range due to a very large bandwidth, and they can operate on high as well as low data rates. Also, physical layers with non-linear rate functions, like 802.11a/g, may operate in the linear regime if the received power is low (e.g. distances between nodes are large). Like in Chapter 5, our findings in this chapter are for networks whose physical layers operates in the linear regime.

6.2.4 Rate Maximization and Power Minimization

There are two typical deployment scenarios for wireless networks: high bit-rate networks and low power consumption networks. The first one considers real-time video and audio communication, web surfing, data transfer, and the like. The primary design focus here is to maximize available rates, subject to power constraints. Typical examples of this type of networks are 802.11 and 802.15.3a wireless LANs and CDMA-HDR cellular systems. We call this case *rate maximization scenario*; here we are interested in the set of feasible rates.

The second scenario is focused on low power networks like sensor networks or networks of computer peripherals. The main goal is to maximize network lifetime, or equivalently, to minimize average consumed power. At the same time, end-to-end flow rates are lower bounded by application requests, and each sender typically has a minimum amount of information to send to a destination in a given time. Here we are interested in minimizing power consumption, subject to minimum long term rate constraints. Long-term average power consumption is defined in Section 6.3.3. We call this case *power minimization scenario*; here we are interested in the set of feasible power allocations.

Different performance objectives for comparing the feasible sets in both scenarios are presented in detail in Section 6.3.4.

6.2.5 Power Control in Existing Systems

The goal of power control is to determine which power a transmitter should use when transmitting a packet. The optimal transmitted power of a packet depends on a large number of

parameters, such as the distance from the destination, the background noise, the amount of interference incurred by concurrent transmissions, etc. In an ad-hoc network, the optimal power also depends on transmitting powers of other concurrently scheduled links. Since power control is tightly coupled with scheduling, it is typically implemented within the MAC protocol.

Perhaps the simplest way to choose the transmitted power is not to do any power control. In other words, whenever a packet is sent, it is sent with maximum allowed power. We call this $0 - P^{MAX}$power control. The $0 - P^{MAX}$power control was widely used in the design of the first wireless MAC protocols, such as 802.11, due to its simplicity, and due to the fact that the optimal power control was not well understood.

Much of the research on power control is focused on voice cellular systems. Those systems typically use quasi-orthogonal channels for different users (e.g. CDMA spreading) in order to decrease multi-user interference, i.e. interference between competing users in the same network. However, the orthogonality of channels is not complete, and some amount of interference between users cannot be avoided (this is captured by the orthogonality factor in Section 6.4.1). Classically, the physical layer of CDMA systems is designed to operate when multi-user interference is small; otherwise (this is is known as the *near-far problem*), signal acquisition and decoding do not work. This is why such systems must employ some form of power control; for example, on the CDMA-HDR uplink, the near-far problem is avoided by equalizing all received powers at the base station. Some pioneering work in this area can be found in [25, 5, 97, 24]. An attempt to design an optimal power control protocol for 802.11 networks has been made in [45, 60, 64, 61]. They consider the 802.11b physical layer with a fixed rate, and the common conclusion is that the power should be adjusted to the minimal value required to be successfully decoded at the destination. For more details on all these protocols, see the related works in Chapter 2.

Several power adaptation protocols have been proposed for power minimization scenarios. A typical example is given in [77] where the power of a link is adjusted to a minimum necessary to reach a destination, and the routing is chosen to minimize the overall power dissipation.

In most of these existing systems, the benefits of power control derive from assumptions on the physical layer (such as fixed rate coding, or the need to avoid near-far problems). It is however possible to do without such assumptions: some examples are the CDMA-HDR downlink (which does rate adaptation), or TH-UWB systems with interference mitigation [58]. This motivates us to pose the problem of optimal MAC design in general terms, assuming power control is an option but not a requirement. Protocols that consider rate adaptation, power adaptation and scheduling in this general setting have been proposed in [19, 21]: they focus on low processing

gain CDMA or UWB networks (thus linear regime) and show that $0 - P^{MAX}$ power control is optimal when the objective is to maximize the total sum of rates. However, as we have shown in Chapter 4, this objective is known to be defective as it requires that the most expensive links be shut down completely. We go beyond these results and establish the optimality of $0 - P^{MAX}$ for any performance objective, and the non-optimality of any non $0 - P^{MAX}$ power control for rate maximization scenarios.

6.2.6 Performance Comparison

For different power control strategies, we are interested in comparing the resulting average rate or power allocations. By using different scheduling strategies with one power control strategy however, one can obtain different average rate or power allocations. The set of all possible rate and power allocations that can be obtained with a given power control strategy, and with different schedules, is called the feasible rates set and the feasible power set, respectively. In this chapter we use the most general way to compare performances of the two power control strategies, which consists in comparing the sets of their Pareto efficient feasible rate and power allocations. Precise definitions of these concepts are given in Section 6.3.4.

6.3 System Assumptions and Modeling

We use the same model of a network as defined in Section 4.3. We analyze an arbitrary multi-hop wireless network that consists of a set of nodes, and every two nodes that directly exchange information are called a link. For each pair of nodes we define a signal attenuation, i.e. a level of signal received at the receiver, assuming the sender is sending with unit power. This attenuation is usually a decreasing function of a link size due to power spreading in all directions, but here we assume it can be an arbitrary number defined for each pair of nodes. Signal attenuation also changes in time due to mobility and different variations of characteristics of paths the signal takes, thus we will model it as a random process.

We model rate as a function $r(\mathrm{SINR})$ of the signal-to-interference-and-noise ratio at the receiver, which is the ratio of received power by the total interference perceived by the receiver including the ambient noise and the transmissions of other links that occur at the same time. We focus on physical layers with a linear rate function. This is the same model as described in Section 5.3.2.

6.3.1 MAC Protocol

The model of the MAC protocol is the same one described in Section 4.3.2. A part of a MAC is a power control strategy. The power control strategy is defined by a set of possible powers that can be allocated to links in any slot. An example of power control strategy is $0 - P^{MAX}$power control where any link in any slot can send with power P^{MAX} or stay idle. This is the simplest strategy where powers are fixed and there is no power adaptation.

Having chosen a power control strategy, a MAC chooses a schedule and assigns powers that belong to the set of possible powers to links in each slot. Finally, the rate on each link in each slot is adapted to the SINRs at receivers.

We assume random fading. Since we have an idealized MAC protocol, it can instantly adapt the schedule and the power and rate allocation to any state of the random fading of links. For a precise mathematical model of MAC protocol, see Section 6.4.2.

6.3.2 Routing Protocol and Traffic Flows

We assume an arbitrary but constant routing protocol (routes and flow demands do not change in time). Flows between sources and destinations are mapped to paths, according to some rules specific to the routing protocol. At one end of the spectrum, nodes do not relay and only one-hop direct paths are possible. At the other end, nodes are willing to relay data for others and multi-hop paths are possible. All these cases correspond to different constraint sets in our model, as defined in Section 6.4.2. We restrict to single-path routing. Sources can send to several destinations (multicast) or to one (unicast).

6.3.3 Power and Rate Constraints

There are four types of power and rate constraints in a wireless network: peak power constraint, short-term average power constraint, long-term average power constraint and average rate constraint. The power constraints are already described in Section 4.3.4. Here we describe only the average rate constraint.

In networks such as sensor or peripheral networks, the goal is to minimize power consumption and to maximize the lifetime of nodes rather than maximize the rates of links. Still, there is a lower bound on the rate a node has to transmit. For example, a temperature sensor on a car engine or a computer mouse have a well-defined rate of information they need to communicate to a central system. This is what we call the *average rate constraint* and we define it as the average

amount of bits a node has to transmit over the network in one second. We assume this average limit is the same on both long and short timescales.

We incorporate explicitly in our model the transmission power constraints, the average consumed power constraints and the average rate constraints. The peak power is incorporated implicitly through the choice of the rate function.

6.3.4 Performance Objectives

Design criteria in wireless networks can be divided into two groups: rate maximization and power minimization. We first consider rate maximization. Given a network topology and a family of MAC protocols, one can define a set of feasible rate allocations as the set of all rate allocations that can be achieved on the network with some MAC protocol from the given family. An interesting subset of the feasible rate set is the set of Pareto efficient rate allocations. A rate allocation is *Pareto efficient* if no rate can be increased without decreasing some other rate. When maximizing rates, we are clearly interested only in the Pareto efficient rate allocations.

The most general way to compare two families of network protocols on a same network is to compare their Pareto efficient rates' sets. If all Pareto efficient rates of one family of protocols are feasible under the other family of protocol, then one can undoubtedly say that the second family is as good as the first. Furthermore, if neither of the Pareto efficient rates of the second family is achievable under the first family of MAC protocols, then we can say that the second family is strictly better than the first. We will use this criterion to compare different power control strategies throughout the chapter. We use the analog approach to compare different power minimization scenarios: in this case a power allocation is Pareto efficient if no average power can be decreased without increasing some other power. Mathematical definitions of terms are given in Section 6.4.2.

We illustrate this comparison on an example: Consider a familly of MAC protocols that use $0 - P^{MAX}$ power control, and an arbitrary scheduling strategy. For a given network topology, each chosen scheduling strategy will give us a different rate allocation, and the set of all rate allocations we can achieve by varying scheduling is the set of feasible rates achieved under $0 - P^{MAX}$ power control. We can do the same experiment with an another power control strategy, obtain its feasible rate set, and then compare the two power control strategies by comparing the Pareto efficient sets as described above.

6.4 Mathematical Model

6.4.1 Notations

We use the same notation as described in Section 4.4.1. We also assume a rate is a linear function of SINR and that there is an orthogonality factor β, both defined in Section 5.3.2.

In addition, we define $\overline{\mathbf{F}}^{MIN} \in \mathbb{R}^I$ to be the vector of minimum average rates achieved by end-to-end flows (every flow may have a different minimum average rate).

6.4.2 Mathematical Formulation

We assume that for every state s there is a schedule consisting of time slots $n = 1...N(s)$ of frequency $\alpha_n(s)$. This is an abstract view of the MAC protocol, without overhead. We normalize these lengths such that $\sum_{n=1}^{N(s)} \alpha_n(s) = 1$. Let us call $\mathbf{p}^n(s)$ the vector of transmission powers assigned to links in slot n and state s, and let $\mathbf{SINR}^n(s)$ be the vector of signal-to-interference-and-noise ratios at receivers of the links, induced by $\mathbf{p}^n(s)$. The rate achievable on link l in slot n and state s is $x_l^n(s) = K \operatorname{SINR}_l^n(s)$. The vector of average rates on the links is thus $\bar{\mathbf{x}} = \mathbb{E}\left[\sum_{n=1}^{N(S)} \alpha_n(S)\mathbf{x}^n(S)\right]$, averaged over the distribution of states. Since $\mathbf{x}^n(s)$ has dimension L (where L is a number of links), by virtue of Carathéodory theorem, when in state s, it is enough to consider $N(s) \leq N = L + 1$ time slots of arbitrary lengths $\boldsymbol{\alpha}(s)$ in order to achieve any point in the convex closure of points $\mathbf{x}^n(s)$.

Feasible rate and power allocations: Given a network topology and a routing matrix R, we define the *set of feasible average powers, link rates and end-to-end rates* $\mathcal{T}$ (without average power or rate constraints). It is the set of $\mathbf{f} \in \mathbb{R}^I$, $\bar{\mathbf{x}} \in \mathbb{R}^L$ and $\bar{\mathbf{p}} \in \mathbb{R}^L$ such that there exist schedules $\boldsymbol{\alpha}(s)$, sets of power allocations $\mathbf{p}^n(s)$ and corresponding sets of rate allocations $\mathbf{x}^n(s)$ for all $n = 1 \cdots N$ and all states $s \in \mathcal{S}$, such that the following set of equalities and inequalities are satisfied for all $n = 1 \cdots N, i = 1 \cdots I, l = 1 \cdots L, o = 1 \cdots O$:

$$\begin{aligned} R\mathbf{f} &\leq \bar{\mathbf{x}} \\ \bar{\mathbf{p}} &= \mathbb{E}\left[\sum_{n=1}^{L+1} \alpha_n(S)\mathbf{p}^n(S)\right] \\ \bar{\mathbf{x}} &= \mathbb{E}\left[\sum_{n=1}^{L+1} \alpha_n(S)\mathbf{x}^n(S)\right] \\ \mathbf{x}_l^n(s) &= K \operatorname{SINR}_l(\mathbf{p}^n(s)) \\ \operatorname{SINR}_l(\mathbf{p}^n(s)) &= \frac{p_l^n(s)h_{ll}(s)}{\eta_l(s)+\beta\sum_{k\neq l} p_k^n(s)h_{kl}(s)} \end{aligned} \tag{6.1}$$

$$\begin{aligned} 1 &= \sum_{n=1}^{L+1} \alpha_n(s) \\ 1 &\geq \sum_{l\,:\,l.\mathrm{src}=o} 1_{\{p_l^n(s)>0\}} + \\ &\quad \sum_{l\,:\,l.\mathrm{dst}=o} 1_{\{p_l^n(s)>0\}} \\ p_l^n(s) &\leq P_l^{MAX} \end{aligned}$$

where $l.\mathrm{src} = o$ and $l.\mathrm{dst} = o$ are true if node o is the source or the destination of link l, respectively.

We are interested in comparing average rates and power consumptions with $0 - P^{MAX}$ and with arbitrary control. With $0 - P^{MAX}$ power control, a node sends with maximum power when sending. More formally this means that in any slot n, power allocation vector $\mathbf{p}^n$ has to belong to the set of extreme power allocations $\mathcal{P}^E = \{\mathbf{p} \,|\, (\forall l = 1 \cdots L)\, p_l \in \{0, P_l^{MAX}\}\}$. In contrast, with an arbitrary power control, any power from the set of all possible power allocations $\mathcal{P}$ is possible. The set $\mathcal{P}$ is defined as $\mathcal{P} = \{\mathbf{p} \,|\, (\forall l = 1 \cdots L)\, p_l \in [0, P_l^{MAX}]\}$.

We say that an average rate allocation $\mathbf{f}$ and average power consumption $\bar{\mathbf{p}}$ is *achievable* with a set of power allocations belonging to $\mathcal{P}$ if for all $n = 1 \cdots N, i = 1 \cdots I, l = 1 \cdots L, o = 1 \cdots O$, it satisfies constraints (6.1), and for all $n = 1 \cdots N, s \in \mathcal{S}, \mathbf{p}^n(s) \in \mathcal{P}$.

We can similarly define the set of average end-to-end rates, link rates and power allocations $\mathcal{T} = \mathcal{T}(\mathcal{P})$ that is achievable with power allocations belonging to $\mathcal{P}$, as the set of all $(\mathbf{f}, \bar{\mathbf{x}}, \bar{\mathbf{p}})$ that are achievable using power allocation $\mathcal{P}$. Thus, sets $\mathcal{T}$ and $\mathcal{T}(\mathcal{P}^E)$ represent the sets of all possible average end-to-end rates, link rates and power consumptions with an arbitrary and with $0 - P^{MAX}$ power control, respectively.

When we consider rate maximization under constraints on average consumed power, we are interested only in the set of feasible rates. If the average consumed power is limited by $\overline{\mathbf{P}}^{MAX}$, then the set of feasible rates is $\mathcal{F} = \{\mathbf{f} \,|\, (\mathbf{f}, \bar{\mathbf{x}}, \bar{\mathbf{p}}) \in \mathcal{T}, \bar{\mathbf{p}} \leq \overline{\mathbf{P}}^{MAX}\}$. Similarly, with $0 - P^{MAX}$ power control, the set of feasible rate is $\mathcal{F}^E = \{\mathbf{f} \,|\, (\mathbf{f}, \bar{\mathbf{x}}, \bar{\mathbf{p}}) \in \mathcal{T}(\mathcal{P}^E), \bar{\mathbf{p}} \leq \overline{\mathbf{P}}^{MAX}\}$. For notational conveience, we analogly define $\bar{\mathcal{X}} = \{\bar{\mathbf{x}} \,|\, (\mathbf{f}, \bar{\mathbf{x}}, \bar{\mathbf{p}}) \in \mathcal{T}, \bar{\mathbf{p}} \leq \overline{\mathbf{P}}^{MAX}\}$ and $\bar{\mathcal{X}}^E = \{\bar{\mathbf{x}} \,|\, (\mathbf{f}, \bar{\mathbf{x}}, \bar{\mathbf{p}}) \in \mathcal{T}(\mathcal{P}^E), \bar{\mathbf{p}} \leq \overline{\mathbf{P}}^{MAX}\}$.

Similarly, when considering power minimization, we focus on the set of feasible average consumed powers. If the average end-to-end flow rate is lower-bounded by $\overline{\mathbf{F}}^{MIN}$, then the set of feasible average consumed powers, under arbitrary power control, is $\overline{\mathcal{P}} = \{\bar{\mathbf{p}} \,|\, (\mathbf{f}, \bar{\mathbf{x}}, \bar{\mathbf{p}}) \in \mathcal{T}, \mathbf{f} \geq \overline{\mathbf{F}}^{MIN}\}$. Similarly, with $0 - P^{MAX}$ power control, the set of feasible rate is $\overline{\mathcal{P}}^E = \{\bar{\mathbf{p}} \,|\, (\mathbf{f}, \bar{\mathbf{x}}, \bar{\mathbf{p}}) \in \mathcal{T}(\mathcal{P}^E), \mathbf{f} \geq \overline{\mathbf{F}}^{MIN}\}$.

Performance Objectives: Finally, we formally define notion of Pareto efficiency that was introduced in Section 6.3.4. Rate vector $\mathbf{f} \in \mathcal{F}$ is Pareto efficient on $\mathcal{F}$ if there exist no other vector $\mathbf{f}' \in \mathcal{F}$ such that for all $i, \mathbf{f}'_i \geq \mathbf{f}_i$ and for some $j, \mathbf{f}'_j > \mathbf{f}_j$. Average power dissipation vector

$\mathbf{p} \in \overline{\mathcal{P}}$ is Pareto efficient on $\overline{\mathcal{P}}$ if there exists no other vector $\mathbf{p}' \in \overline{\mathcal{P}}$ such that for all $i, \mathbf{p}'_i \leq \mathbf{p}_i$ and for some $j, \mathbf{p}'_j < \mathbf{p}_j$.

6.5 Main Findings

6.5.1 Rate Maximization

In this subsection we show that any rate allocation that is feasible with an arbitrary power control and under some average power constraint, is also achievable with $0 - P^{MAX}$power control. Moreover, if we consider a scenario without average power constraints, then $0 - P^{MAX}$is the only optimal power control.

We first show how an arbitrary schedule and power allocation can be improved simply by using $0 - P^{MAX}$approach. Consider an arbitrary power allocation $\mathbf{p}$ and schedule $\boldsymbol{\alpha}$, and the resulting average link rate allocation $\bar{\mathbf{x}}$, power dissipation $\bar{\mathbf{p}}$ and flow rate allocation .

Theorem 6.1 *Suppose that in some slot n, for some link l we have $0 < p_l^n < P_l^{MAX}$. Then we can construct a new schedule by dividing slot n into two subslots n_1 and n_2, with $p_l^{n_1} = P_l^{MAX}$ and $p_l^{n_2} = 0$ such that we do not decrease any component of $\bar{\mathbf{x}}$. Furthermore, if there are no constraints on average consumed power (or equivalently $\overline{\mathbf{P}}^{MAX} \geq \mathbf{P}^{MAX}$), we can increase at least one component of $\bar{\mathbf{x}}$ without decreasing any other component. Conversely, we can construct a new schedule in the same manner such that we do not increase any of components of $\bar{\mathbf{p}}$.*

Proof: The basic idea of the proof is given in [6]. We formalize it here using our network model. We divide slot n in two slots, n_1 and n_2, such that $\alpha_{n_1} P_l^{MAX} = \alpha_n p_l^n$ and $\alpha_{n_2} = \alpha_n - \alpha_{n_1}$. In slot n_1 we let link l transmit with the full power P_l^{MAX} and in the second slot we shut it down ($p_l^{n_2} = 0$). We do not modify other links or other slots.

For some $k \neq l$ and if $p_k^n > 0$, it is easy to verify that rate $x_k^n(p_l^n)$ is a strictly concave function. Therefore, we have $x_k^n(p_l^n) < \alpha_{n_1} x_k^{n_1}(P_l^{MAX})$. Also, the rate of link l stays the same as $\alpha_n x_l^n = \alpha_{n_1} x_l^{n_1} + \alpha_{n_2} x_l^{n_2}$. We conclude that if some $k \neq l$ we have $p_k^n > 0$, we increase the rate of at least one compontent of $\bar{\mathbf{x}}$ without decreasing rates of others.

If on the contrary for all $k \neq l$ we have $p_k^n = 0$, then link l does not interfere with anyone, and we do not decrease any rate. Furthermore, if $\overline{\mathbf{P}}^{MAX} \geq \mathbf{P}^{MAX}$, then we can set $\alpha_{n_1} = \alpha_n$. This way we increase the rate $\bar{\mathbf{x}}_l$ without decreasing other rates.

Similar reasoning holds when considering $\bar{\mathbf{p}}$. Here again if for all $k \neq l$ we have $p_k^n = 0$, we cannot decrease $\bar{\mathbf{p}}_l$ whatever kind of rescheduling we do during slot n. Therefore, we can only guarantee that the proposed rescheduling will not increase power dissipation.

q.e.d.

Next we want to show $\mathcal{F}^E = \mathcal{F}$. We clearly have $\mathcal{F}^E \subseteq \mathcal{F}$, and we want to show that every feasible flow rate allocation can be achieved by a set of extreme power allocation from $\mathcal{P}^E$, that is $\mathcal{F} \subseteq \mathcal{F}^E$. In other words that every feasible flow rate allocation can be achieved only with an appropriate scheduling, and without power control.

Theorem 6.2 *If the rate is a linear function of the SINR, then for arbitrary values of parameters of constraint set (6.1), we have that $\mathcal{F}^E = \mathcal{F}$.*

Proof: Proof of the theorem follows directly from Theorem 6.1. Any feasible $\bar{\mathbf{x}}$, hence any feasible can be achieved starting from an arbitrary schedule and modifying each slot that contains a non-$0 - P^{MAX}$power allocation as described in the proof of Theorem 6.1.

q.e.d.

Theorem Theorem 6.2 says that every feasible rate allocation, thus including the Pareto efficient ones, can be achieved with $0 - P^{MAX}$power control. Hence $0 - P^{MAX}$is at least as good as any other power control, and power adaptation is not needed.

To interpret this finding, consider a UWB MAC protocol presented in [21] where both power adaptation and scheduling is used. Any rate achieved by this MAC protocol could be achieved with another protocol that would not adapt power and would use an appropriate scheduling.

We next consider a scenario where there are no constraints on average consumed power (or equivalently $\overline{\mathbf{P}}^{MAX} \geq \mathbf{P}^{MAX}$), and we prove that power adaptation is strictly suboptimal.

Theorem 6.3 *Consider an arbitrary network where the rate is a linear function of the SINR, and an arbitrary schedule $\boldsymbol{\alpha}$ and a set of power allocations $\mathbf{p}^n$ for that network. If for some n, $\alpha^n > 0$ and power allocation $\mathbf{p}^n \notin \mathcal{P}^E$ then the resulting average rate allocation $\mathbf{f}$ is not Pareto efficient on $\mathcal{F}$.*

To prove the theorem we need the following lemma:

Lemma 6.1 *Let $(\mathbf{f}, \bar{\mathbf{x}}, \bar{\mathbf{p}}) \in \mathcal{T}$. If $\mathbf{f}$ is Pareto efficient on $\mathcal{F}$ then the corresponding average link rate $\bar{\mathbf{x}}$ has to be Pareto efficient on $\bar{\mathcal{X}}$.*

Proof: We proceed by contradiction. Suppose that $\mathbf{f}$ is Pareto efficient on $\mathcal{F}$ but $\bar{\mathbf{x}}$ is not Pareto efficient on $\bar{\mathcal{X}}$, and we can increase $\bar{x}_i$ for some i without decreasing other rates. In other words, we have $(\mathbf{f}, \bar{\mathbf{x}}', \bar{\mathbf{p}}) \in \mathcal{T}$ such that $\bar{x}'_i = \bar{x}_i + \epsilon$ and $\bar{x}'_k = \bar{x}_k$ for all $k \neq i$. If there is flow f_j such that link i is its bottleneck, than we can increase the rate of f_j since $\bar{x}'_i > \bar{x}_i$, hence $\mathbf{f}$ is not Pareto efficient. Hence we conclude no flow has a bottleneck on link i.

We next choose an arbitrary flow k, we start from link rate allocation $\bar{\mathbf{x}}'$ and we show how to construct a schedule that will increase the rate of flow k. Let us denote with j the bottleneck link of flow k (if k has no bottleneck $\mathbf{f}$ is obviously not Pareto efficient). We first try to find a slot in which both j and i are active. If this slot exists (say it is slot n) than we can decrease the power $p_i^n(s)$ (in some state s with positive probability) by some $\epsilon_i > 0$ such that the resulting $\bar{x}_i''$ has the property $\bar{x}_i' > \bar{x}_i'' > \bar{x}_i$. Since link j is also active in slot n and we have decrease the interference (due to the assumption, $h_{ij}(s) > 0$), we have also increased $\bar{x}_j'' > \bar{x}_j$, and we can in turn increase f_k which violates the Pareto efficient property of $\mathbf{f}$.

Finally we have the case when links i and j are not active in the same slot. We pick slot n_i and n_j, in which i and j are active, respectively, such that $\alpha_i^n(s) > 0$ and $\alpha_j^n(s) > 0$ (for some states s with positive probability). We then decrease the power $p_i^{n_i}$ by some $\epsilon_i > 0$ such that the resulting $\bar{x}_i''$ has the property $\bar{x}_i' > \bar{x}_i'' > \bar{x}_i$. At the same time, we increased the average rates of all links scheduled during the slot $\alpha_i^n(s)$. Now we can decrease the duration of the slot $\alpha_i^n(s)$ such that in the new allocation all those links will have at least the same rates as in the initial configuration $\bar{x}_l'' \geq \bar{x}_l$. However, since we decreased $\alpha_i^n(s)$, we can now increase $\alpha_j^n(s)$, hence also increase $\bar{x}_j$. Now flow k looses its bottleneck, hence we can increase f_k again violating Pareto efficient assumption.

q.e.d.

Proof of Theorem 6.3: We proceed by contradiction, and assume there exist a schedule $\boldsymbol{\alpha}$ and a set of power allocations $\{\mathbf{p}^n(s)\}_{1,\cdots,N}$ such that the resulting average rate allocation $\mathbf{f}$ is Pareto efficient (and thus on the boundary of set $\mathcal{F}$), and for some n, i, $0 < \mathbf{p}_i^n < P_i^{MAX}$. From Lemma 6.1 we have that since $\mathbf{f}$ is Pareto efficient on $\mathcal{F}$ then is so $\bar{\mathbf{x}}$ on $\bar{\mathcal{X}}$. Now, according to Theorem 6.1, since there are no average power constraints, we can reschedule slot n such that we increase at least one link rate, hence $\bar{\mathbf{x}}$ is not Pareto efficient on $\bar{\mathcal{X}}$ that yields to contradiction.

q.e.d.

The theorem says that a Pareto efficient allocation cannot be achieved if in any time slot a power allocation different from $0 - P^{MAX}$is used. Applying the finding on the framework of [21] we can conclude that there exists a different schedule that does not use power control and that improves the performance of a network.

The main use of Theorem 6.2 and Theorem 6.3 is the following corollary:

Corollary 6.1 *When the rate is a linear function of the SINR,* $0-P^{MAX}$*power control is actually the single optimal power control strategy, and any other power adaptation is strictly suboptimal.*

6.5.2 Power Minimization

We next analyze the effect of power adaptation on minimizing dissipated power of a network. By decreasing transmitting power, one decreases the dissipated power and also the destructive effect of interference on others, hence, intuitively, power control should minimize power consumption. However, as we show here, in the case of linear rate function, power control does not bring any benefit.

Theorem 6.4 *If the rate is a linear function of the SINR, then for arbitrary values of parameters of the constraint set (6.1), we have that* $\overline{\mathcal{P}}^E = \overline{\mathcal{P}}$.

Consider an arbitrary $\bar{\mathbf{p}} \in \overline{\mathcal{P}}$. By virtue of Theorem 6.1 we can always construct a new schedule that utilizes only $0 - P^{MAX}$ power allocations and achieves the same $\bar{\mathbf{p}}$.

q.e.d.

All feasible average power dissipations, hence all Pareto efficient ones, can be achieved with $0 - P^{MAX}$ power control, hence it is at least as good as any other power control. Again here, power adaptation is not needed. We note here that for power minimization there is no statement analog to Theorem 6.3. Theorem 6.3 assumes that there are no average power constraints. In the framework of power minimization, this corresponds to a setting with no average rate constraints, which leads to the trivial solution of having the network silent all the time.

6.5.3 Numerical Example

In order to illustrate the above findings we give a simple example. Consider a network of two links presented on the left of Figure 6.1. This network is known as the near-far scenario as an interferer is closer to a receiver than the corresponding transmitter. Node S_1 transmits to D_1 and node S_2 transmits to D_2. We introduce two simple MAC protocols. The first MAC protocol assumes $0 - P^{MAX}$ power control and arbitrary scheduling. The second assumes no scheduling (constant power allocations through time, like in some cellular systems), and arbitrary power control strategy. The corresponding sets of feasible rates and powers are given on the right of Figure 6.1.

We see that when maximizing rates, only $0 - P^{MAX}$ power control can achieve all feasible rates, including the Pareto efficient ones. On the contrary, the second MAC protocol that does not use scheduling but uses power adaptation achieves only a fraction of feasible rates. Furthermore, only in cases when power allocation is $0 - P^{MAX}$, the achieved rates are Pareto efficient.

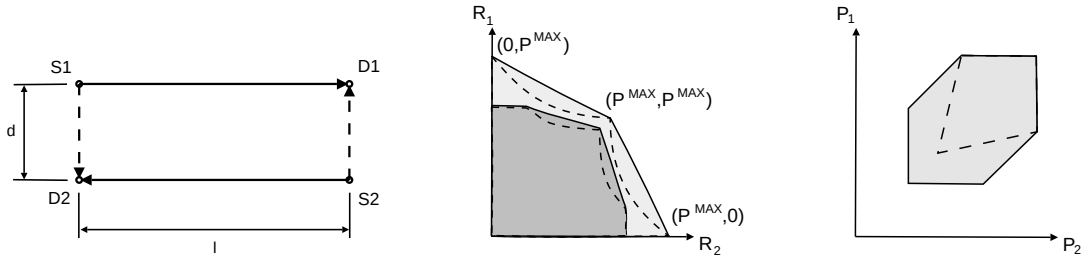


Figure 6.1: A simple example of a network with 2 links. The topology of the network is given on the left. Node S_1 sends to node D_1 while node S_2 sends to D_2. The feasible rate set for this network is given in the middle. The lighter region in dashed lines represents the set of feasible rates that can be achieved without scheduling, only with power adaptation. The lighter region in full lines represents an increase that is achieved by scheduling and without power adaptation ($0 - P^{MAX}$power control). The darker region in dashed lines is the same example without scheduling and with power control, but this time with additional average power constraints. Again the darker region in full lines represents an improvement introduced by scheduling. We see that the second protocol cannot achieve Pareto efficient rates of the feasible rate set, except for the three rate allocations. But these three rate allocations are achieved with power allocations $(0, P^{MAX})$, $(P^{MAX}, 0)$ and (P^{MAX}, P^{MAX}) which belong to $0 - P^{MAX}$power strategy. In the figure on the left, the feasible set of average consumed power under minimum rate constraints is depicted in gray. The region in full lines represents average power consumption achievable with scheduling and without power adaptation, and the region in dashed lines represents average power consumptions achievable without scheduling and with power adaptation. All average powers belonging to this set can be achieved without power adaptation.

However, when there is an average power limit, there might exist a schedule and a power control strategy, different from $0 - P^{MAX}$, that can achieve Pareto efficient allocations, as discussed in Section 6.5.2. To see this, consider an even simpler example of a single link. Let P^{MAX} be the maximum transmitting power, $\overline{P}^{MAX} < P^{MAX}$ the maximum average consumed power, h be the fading from the source and η be the power of background white noise. There exist only one Pareto efficient rate allocation which is $R = \overline{P}^{MAX} h/\eta$. It can be achieved by sending $\alpha = \overline{P}^{MAX}/P^{MAX}$ fraction of the time using full power, or by sending all the time using $\overline{P}^{MAX}$ as the transmitting power. The second strategy thus does not have the form of $0 - P^{MAX}$ power control, yet it achieves the Pareto efficient allocation. An analogous construction can be made to show that a non-$0 - P^{MAX}$ power control can achieve Pareto efficient average power allocation.

On the right of Figure 6.1 we depict the feasible average dissipated powers allocations for an arbitrary power control, and for $0 - P^{MAX}$power control. We see that the two sets coincide.

6.6 Summary of Main Findings and Conclusions

We consider a general multi-hop wireless network with random channels (due to fading or mobility), where link rates, transmission powers and medium access can be varied, and we focus on physical layers that operate in the linear regime. For such systems, one can find rate control, power control and theoretical MAC protocols that maximize the performance. This is a joint optimization problem and a change in any of the three components influences the choice of the other two. We consider different power control strategies, for each of them we assume the optimal MAC and rate adaptation, and we compare their performances. The goal is to characterize the optimal power control. Our main findings are:

Any feasible rate allocation can be achieved with $0 - P^{MAX}$ - We consider the rate maximization scenario and we mathematically prove that every feasible rate allocation can be achieved without power control (power adaptation is not needed beyond $0 - P^{MAX}$).

Pareto efficient rate allocations can only be achieved with $0 - P^{MAX}$ - We again consider rate maximization scenario and we prove that, if there are no average power constraints (i.e. only peak power constraints), any power control that does not use $0 - P^{MAX}$ power control is not Pareto efficient (power adaptation is suboptimal).

Any feasible average power dissipation can be achieved with $0 - P^{MAX}$ - We further consider the power minimization scenario. We prove that any feasible average power allocation is achievable without power adaptation. In other words, any feasible average power allocation is achievable with $0 - P^{MAX}$ power control and an appropriate schedule, and power adaptation is not needed.

The above findings are based on the observation that any schedule can be modified to use only $0 - P^{MAX}$, such that the resulting link rates and average power dissipations improve, or stay the same.

Our results do suggest that, for multihop networks operating in the linear regime and that can live with arbitrary levels of rate and power, power control beyond $0 - P^{MAX}$ can be avoided, and thus, the MAC layer should concentrate on scheduling (by means of a protocol) and rate adaptation, using full power whenever a transmission is allowed by the protocol.

If the number of possible physical link rates is small, power adaptation and scheduling should be used (as for example in [61]), but if the number of possible link rates is large, which is usually the case with adaptive modulation and/or coding, the rates should be adapted using $0 - P^{MAX}$ power control and scheduling.

Another conclusion that stems from our work is that, unlike the common belief that for rate maximization, in CDMA or similar data networks with almost-orthogonal links' transmissions, it is better to solve near-far problems by scheduling and rate adaptation and to use $0 - P^{MAX}$ power control, instead of using power adaptation that tends to equalize received powers.

It remains as a future work to further investigate the trade-off between scheduling and power adaptation by incorporating costs of different power control and scheduling protocols.

Chapter 7

Cross Layer Design for Narrow-Band Networks

7.1 Introduction

In the previous chapters, we have analyzed a cross layer design of wide-band wireless networks. We have presented several conclusions on the optimal design, but these conclusions are highly dependent on the assumption that the physical layer is wide-band and the rate function is linear. It is not clear whether the same results will hold in narrow-band counterparts.

In this chapter, we are interested in finding a jointly optimal scheduling, routing rate control and power control that achieve the proportionally fair rate allocation in a narrow-band multi-hop wireless network. This is again a highly complex optimization problem and it has been previously solved but only for small networks. In this chapter, we use different heuristics to obtain an approximate solution to this optimization problem and discover guidelines for an efficient protocol design.

We model a link between two nodes as a point-to-point single user narrow-band Gaussian channel with deterministic fading and a unique power limit for all nodes. This model closely approximates CDMA and 802.11 networks. In the second part, when we analyze realistic examples of networks on a plane, we model the performance of an existing 802.11a/b wireless card [3].

Like in Chapter 5, we first focus on one-dimensional network topologies, where all nodes are aligned on a straight line. These topologies represent a large class of existing networks, from car networks on highways to networks on coasts or mountain valleys. If a network has a finite size we replace a line with a ring in order to avoid border effects.

We also assume that a network is symmetric, that is, all nodes are equally spaced and each

node has data to send to its d hop away neighbor on the right. This simplification allows us to immediately identify the optimal schedule. Although the symmetry argument is not always realistic, it facilitates the analysis and gives us important insight into design of homogeneous networks.

We allow the two most common routing policies. One is that each node sends data directly to its destination (DIR), hence there is no relaying. The other is that each node forwards data to the destination by relaying it over all intermediate nodes, thus using the minimum energy path (MER). There are clearly other relay routing policies that might perform better than minimum energy routing. However, minimum energy routing is frequently used, simple to implement, and performs comparably to direct routing.

We solve the optimization problem by analysis and simulations, as in Chapter 5. We find that for smaller transmission powers it is optimal to relay over other nodes, whereas for high powers it is optimal to send data directly to a destination. We also show when this transition occurs. This is in contrast with the wide-band design, where it is always optimal to relay.

We analyze the optimal schedule and find that if a node is active, it should send at the maximum power, much like in wide-band networks. When transmission power is small, the optimal schedule is that every second node is sending, and as the power grows, the distance between the active nodes grows. Furthermore, in large networks the distance between the nodes sending at the same time is never larger than 4.5 times the size of the links used (number of nodes spanned by one transmission link), and it converges to that value for large transmission powers. This is again in contrast with wide-band design where the distance between active nodes constantly grows with transmission power.

Also, the spacing between active nodes depends on link states, hence on the choice of a routing protocol. This is again in contrast with the wide-band design principles.

We next analyze random networks on a plane, where nodes are uniformly distributed on a square, and each source talks to a randomly chosen destination. Here, we take a 802.11a/g physical layer model, based on an existing 802.11a/b wireless card [3]. Like in Chapter 5, we solve this optimization problem by heuristics that are constructed by using the findings from the linear case.

We show by simulations that the optimal protocol outperforms the existing 802.11 rate adaptation protocols; the exclusion region of 802.11a/g is too large and the spatial reuse is too low, in other words, the efficiency of 802.11 could be improved by allowing more interference.

The above findings give insights into the optimal design of narrow-band networks. They are not thorough, but they suggest the narrow-band design is significantly different from the wide-

band design. They also suggest that the designs of existing protocols for narrow-band multi-hop networks are not optimal, and give directions on how these protocols can be improved.

This chapter is organized as follows: In Section 7.2 we explain the model of a network that we shall use in this chapter. In Section 7.3 we analyze the ring and line models. In Section 7.4 we discuss random networks on a plane. In Section 7.5 we conclude and give directions for future work.

7.2 Assumptions and Modeling

In this chapter, we will use almost the same model for PHY, MAC and routing as in the wide-band case, in Chapter 5. The model is described in detail in Section 5.3. The only difference in the model in this chapter is the rate function, which is described below.

7.2.1 Physical Layer

Similarly to Chapter 5, we assume data is separately encoded on each hop, and we assume that a receiver cannot decode third party communications, hence it treats it as noise. The total noise at a receiver comprises the interference from concurrent transmissions, which are assumed Gaussian, and the white noise. Then each hop can be modeled as a single user Gaussian channel. The capacity of a Gaussian channel can be expressed as [18]

$$\mathbf{x}_i^n = \frac{1}{2}\log_2(1 + \mathbf{SINR}_i^n). \tag{7.1}$$

The Shannon capacity bound for this type of channel can be achieved using contemporary CDMA technology. We will use the exact expression (7.1) as a rate function in the analysis of the ring and line cases.

In the second part, when we analyze networks in plane, we will use the 802.11a/b rate function given in [3]. This function is a stair function which can achieve a certain fixed number of rates, with different given modulation and coding scheme. With a more fine-grained coding the number of achievable rates can be increased, and we interpolate the function between these points to obtain a smooth rate function. The smoothed rate function can then be closely approximated by Equation (7.1). We have used the actual values of rates and SNRs from [3] for the second part of the analysis, in order to be able to present realistic values of rates and distances achieved in the example. We also used realistic values of power constraints for 802.11 networks, which typically vary between 1mW and 100mW.

7.2.2 Optimization Problem

The goal of the problem is to find the proportionally fair rate allocation on the set of feasible rates. It is equivalent to the following optimization problem:

$$U = \max_{R \in \mathcal{R}, \mathbf{f} \in \mathcal{F}(R)} \sum_{i=1}^{I} \log(f_i) \tag{7.2}$$

where $\mathcal{F}(R)$ is the feasible rate set, given routing R, defined as in Equation (5.5), except for the modified rate functions, described in Section 7.2.1. $\mathcal{R}$ is the set of possible routing algorithms defined by Section 5.4.3.

7.2.3 Network Topologies

Ring and Line Case

Finding the optimal solution for the model in (7.2) is a very complex problem in the general case (discussion for arbitrary networks with up to 6 nodes are given in [88]). In order to analyze networks with a larger number of nodes, we first restrict our attention to ring and line topologies with a high level of symmetry, depicted on Figure 7.1.

Figure 7.1: Analyzed topologies: ring and line.

Ring topology represents an L-sided regular polygon with distance l between nodes. It is described in Section 5.5. Line topology is a limiting case when the number of nodes in a ring tends to infinity while l remains fixed.

Networks in Plane

In the second part, we consider a random, two-dimensional network, with 40 nodes distributed on a square of 200m x 200m. Each source randomly chooses its destination.

7.3 The Ring Case

7.3.1 Analysis

In this section we present several analytical findings for our ring model.

Proportional Fairness and Scheduling

We first recall two properties of the ring case that were demonstrated in Chapter 5. These properties do not depend on the nature of the physical layer, and the proofs presented in Chapter 5 hold both in the wide-band and the narrow-band case.

The first property is that in the above defined ring wireless network, if a rate allocation is proportionally fair then each flow has the same rate. This is formulated as Proposition 5.2 in Chapter 5.

The second property is a property of the optimal schedule. It says that the optimal schedule consists of L rotationally symmetric power vectors that are equally frequent. This is formulated as Proposition 5.3 in Chapter 5.

Power Allocation

We explained above that the optimal scheduling strategy is to allocate equal time to each rotation of a single power allocation vector $\mathbf{p}$. We now want to characterize the optimal allocation $\mathbf{p}$. Since the rate function in this case is not linear, the findings about the optimal power control, presented in Chapter 6, do not hold.

We noticed that for an arbitrary feasible power allocation $\mathbf{q}$, if we fix $q_2, \cdots, q_n$, the rate is maximized if q_1 is either 0 or P. We were not able to formally prove this statement. However, we found empirically that the rate is a quasi-convex function [80] of the power q_1, which implies the $0 - P^{MAX}$property:

Proposition 7.1 *Let us consider $x(q_1)$ as a function only of the first component of vector $\mathbf{q}$. Then, for arbitrary values of components $q_2, \cdots, q_n$ of $\mathbf{q}$, function $x(q_1)$ is quasi-convex.*

One way to prove that $x(q_1)$ quasi-convex is to show that if $\frac{\partial x}{\partial q_1} = 0$ then necessarily $\frac{\partial^2 x}{\partial q_1^2} \geq 0$. We numerically tested this claim for rings with L up to 6. We also observed that for large L, when a ring can be approximated with a line, only a few of the closest neighbors significantly contribute to the interference. Specifically, if we take the most dense power allocation where every second node is sending at maximum power, and we take $\gamma = 4$, then nodes further than 3 hops away from a destination contribute to the overall interference 0.47%, and for $\gamma = 2$, it

contributes about 15%. Using this approximation we numerically verified this claim for large n as well.

Proposition 7.2 *If Proposition 7.1 is true, then the optimal power allocation* $\mathbf{p}$ *is* $0 - P^{MAX}$*, that is, each power* p_i *is either 0 or* P^{MAX}*.*

Since Proposition 7.1 is true for arbitrary $q_2, \cdots, q_n$, it is also valid for $p_2, \cdots, p_n$, hence we conclude p_1 in the optimal allocation $\mathbf{p}$ has to be either 0 or P^{MAX}. The same reasoning applies to all coordinates of $\mathbf{p}$.

q.e.d.

Restricting ourselves to power allocations where all nodes either send at the full power or not send at all, we have the following claim:

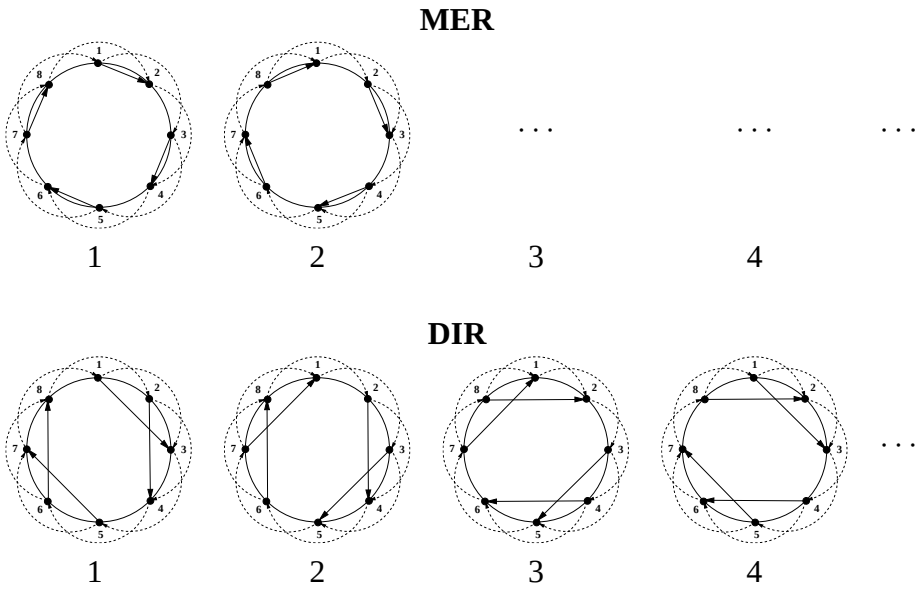
Proposition 7.3 *Optimal power allocation* $\mathbf{p}$ *consists of several groups of adjacent active nodes. Each group has a size from 1 to* d*, and the differences in sizes of any two groups is at most 1. The difference in distances between two adjacent groups of active nodes for any two pairs of groups is at most 1. Especially, in case of MER routing, the size of all active groups is one.*

An example of the optimal power allocation is illustrated in Figure 7.2 and Figure 7.3. We numerically tested this claim and found it true for networks of up to 20 nodes. According to Proposition 7.3 we can then define t as an average size of a group of active nodes, and s as an average distance between the first active nodes of two adjacent groups. Since these parameters vary for at most one, they depict well the optimal power allocation, especially in large networks, and thus in the line case. Note that for MER routing t will always be 1, and for DIR we have $t \leq d$.

Proposition 7.3 is similar to Proposition 5.6 from Chapter 5, except the latter is formulated only for MER routing, and is a special case of the former.

7.3.2 Results

In this section we present numerical solutions to the joint optimization problems for various values of parameters. These parameters are a number of nodes (links) L, a distance to the destination d, a relative maximum power $P^{MAX}K_a l^{-\gamma}/\eta$ and a fading coefficient γ. Again, the two main questions that we seek to answer are when it is beneficial to relay (i.e., when is MER routing better than DIR), and what is the optimal schedule and power allocation, for various values of system parameters.



$n = 8$, $d = 2$, $P = 10^{-2}$

Figure 7.2: Illustration of Proposition 7.3. On the top, the optimal scheduling for the two routing policies is given, for $n = 8$, $d = 2$, $P = 10^{-2}$. Dashed arrows depict flows, and solid arrows depict active links. Numbers below are time slot numbers. Each link is either inactive, or active at the full power. There are 8 allocations in total, but the first 4 are exactly the same as the last 4, for both routing policies. These 8 allocation repeat in a row, each taking the equal time slot, and each one is a rotation of the previous one. In MER routing, distance s between active nodes is 2. In DIR routing, there are 2 groups of $t = 2$ active nodes and distance between them is $s = 4$.

$n = 8$, $d = 2$, $P = 10^{-2}$

MER	DIR
10101010	11001100

$n = 18$, $d = 2$, $P = 10^{-2}$

MER	DIR
101010101010101010	110011001100011000

Figure 7.3: Short representations of the optimal policies for $n = 8$ and $n = 18$: one rotation of the optimal power allocation is given, each link denoted with 0 if inactive and 1 if active. The left case $n = 8$ is the one depicted on Figure 7.2. In MER routing, every second node is active. In DIR routing, for $n = 8$ there are 2 groups of 2 active nodes, and distance between them is 4. For $n = 18$, there are 4 groups of 2 active nodes and distances between them are 4 and 5.

In order to obtain numerical solutions we used different techniques. We first solved the problem over all possible power allocations. Since the rate function is $\log(1 + \mathrm{SINR})$, it is

easy to show that this optimization problem is a d.c. programming problem [71]. Although we solved this problem with a branch and bound approach, the solution was too complex to be applied for networks larger than a few nodes. These solutions however verified Proposition 7.1. We next used Proposition 7.1 and we solved the problem by searching over 2^L possible power allocations for networks with up to 20 nodes. Finally, we used Proposition 7.3 to solve the problem for a line and an arbitrary large ring (since a very large ring can be approximated with a line). In the following sections, we present numerical results and conclusions. Where it is not explicitly stated, we assume $\gamma = 4$. For a discussion on how to choose realistic values of system parameters, see Section 5.7.2.

The Ring Case

We first consider an $n = 18$ node ring. We search over 2^L possible power allocations to find the optimal one for both MER and DIR routing. The results are depicted on Figure 7.4. On the left we see the rate per flow, and on the right we see the rate per distance per flow, as defined in [30]. For small enough powers it is better to use minimum energy routing, hence to relay, and for large power it is optimal to use direct routing.

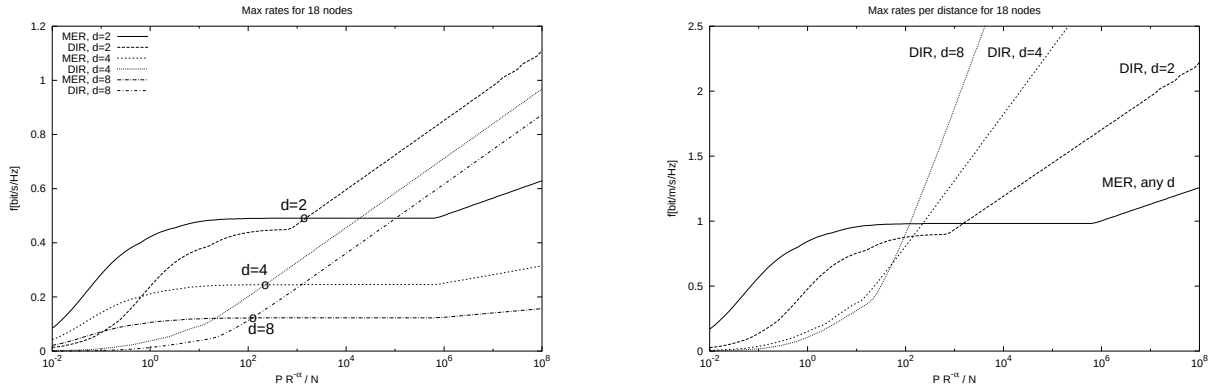


Figure 7.4: On the left, maximum rates for DIR and MER routing and different values of d are depicted, for a ring of 18 nodes. We see that for small transmitting powers MER routing is better then DIR, and for large powers DIR becomes better. The transition power depends on d. On the right, maximum rate per distance is depicted. We see that rate per distance is the same for MER, regardless of d. This is not the case for DIR.

Figure 7.5 shows at what power limit the transition from MER to DIR occurs, for different network sizes and flow lengths. The larger the network is and the shorter the flows are, the more spacial reuse and more incentive to relay there is. We also see that transition power is log-linear with respect to the size of the network L and is super log-linear with respect to the flow length d.

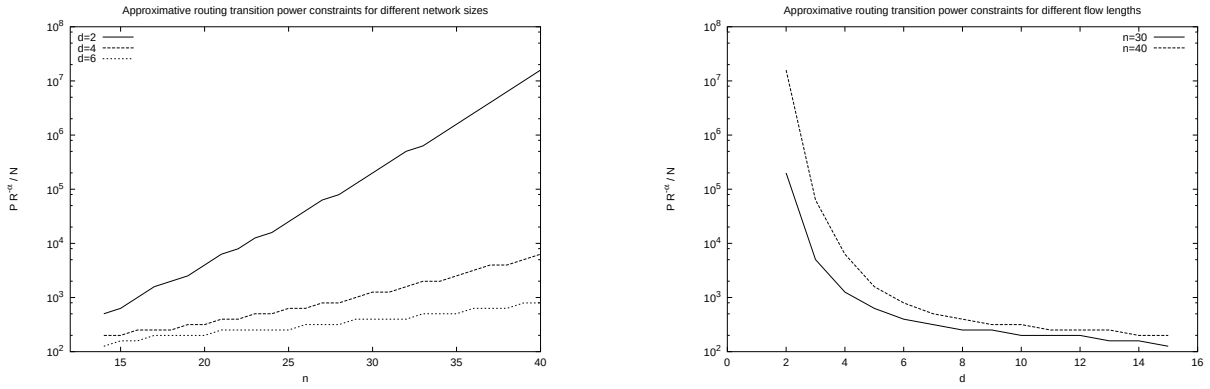


Figure 7.5: The two figures depicts at which power constraint a transition from minimal-energy routing to direct routing occurs, as a function of n and d.

P^{MAX}	MER	s	DIR	(s,t)	Better
10^{-2}	**1010101010101010**	2	111000111000111000	(6,3)	MER
10^{-1}	**1001001001001001000**	3	110000110000110000	(6,2)	MER
1	**1001001000100010000**	3.6	100000000100000000	(9,1)	MER
10	**1000100010000100000**	4.5	100000000100000000	(9,1)	MER
10^5	100010001000010000	4.5	**1000000000000000000**	(18,1)	DIR
10^6	100000000000000000	18	**1000000000000000000**	(18,1)	DIR

Figure 7.6: Optimal scheduling for different power constraints, for both MER and DIR (L=18, d=3). The first column gives the power limit. The second gives the optimal schedule for MER routing, where 1 means a node is on and 0 means a node is off, and s is the average distance. The third column gives the optimal schedule for DIR and the average distance between adjacent active groups s and average size of an active group t(t is always 1 for MER). The last column tells which routing performs better. The optimal joint scheduling and routing policy is shown in boldface.

Another interesting question to consider is the optimal schedule, given the system parameters. An example is given on Figure 7.6, for an 18 node ring with flow length of 3. We see that for very small powers, as many nodes as possible are active at the same time, because the interference is small. In MER this means every second node, hence $s = 2$. A group cannot be larger than one, since link size is 1, hence $t = 1$. In DIR, it means first d nodes are active and the next d nodes are receiving, hence inactive, and so on, leading to $t = d$ and $s = 2d$. When power grows, the distance between active groups s grows, and group size t shrinks. Eventually, for very large powers, we will have only one active node at a time for both routing policies. We then have a log-linear increase of rate with power in both policies, since now power allocation, scheduling and topology are fixed and a change in power limit directly increases rate. It is also interesting to

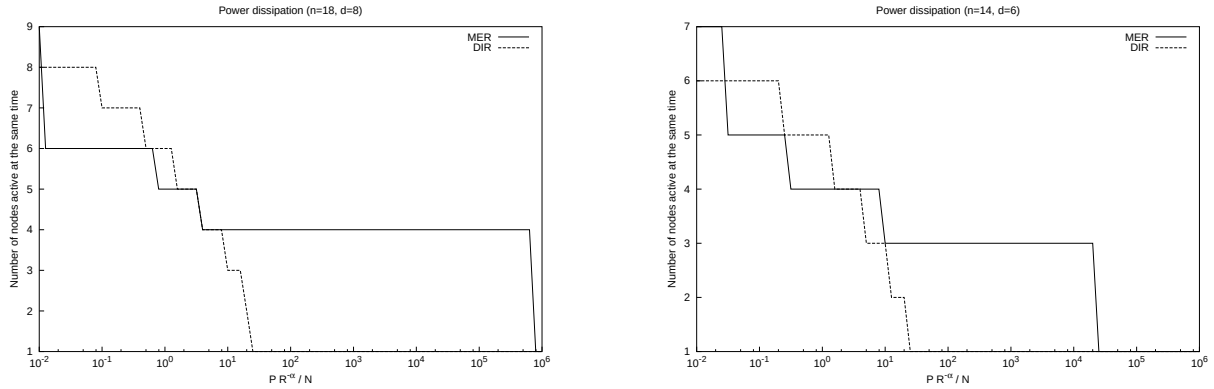


Figure 7.7: Number of nodes active in one time slot. Since the emitting powers or all active nodes are the same, and the number of active nodes is the same in all slot, this number is directly proportional to dissipated energy.

notice that DIR is never better than MER when $t > 1$ since the interfering node is closer to the destination than the sender.

It is interesting to compare this finding with the findings about the exclusion regions from Chapter 5. Consider the example from Figure 7.6. It is easy to verify that in all cases, in the optimal schedule, there are indeed exclusion regions enforced around each active receiver. In case of MER routing, a receiver is right next to a sender. The size of the exclusion region corresponds approximately to $s - 1$ (since a receiver is always located next to a transmitter). In case of DIR routing, we can distinguish 2 cases. The first one is when $t > 1$, which corresponds to the first two lines on Figure 7.6. For $P^{MAX} = 10^{-2}$ we have every node is either sending or receiving, hence the exclusion region is smaller than l. For $P^{MAX} = 10^{-1}$, the exclusion region size is between l and $2l$. All the other lines correspond to the second case, which has $t = 1$, and the exclusion region has the radius of approximately $s - d$.

The number of active nodes can be seen on Figure 7.7. This number is the same for all slots, and since all active nodes use the same power, it is directly proportional to the dissipated energy. We see that for smaller powers, both routing policies dissipate approximately the same energy. When the power gets larger, DIR uses only 1 active node while MER uses 4, hence MER dissipates 4 times more energy at a lower rate (see Figure 7.4). For even larger powers, energy dissipation is the same since only one node is active at a time.

The Line Case

Next, we consider an infinite line with nodes equally spaced on a distance l. Since this is the limiting case of rings when L tends to infinity and distance l between nodes remains constant,

these results also apply to rings with a large L. On the left of Figure 7.8 we see maximum rates for the two routing policies. We first note that DIR routing is never better than MER (it is actually only slightly better, and the reason for this lies in the optimal power allocation, as is discussed later). This is in accordance with Figure 7.5, where we see that a transition power constraint grows exponentially with a number of nodes. From Figure 7.8 we see that for smaller rates, MER performs better, and when power constraint is sufficiently high, DIR and MER performs the same. The same is observed for different values of γ. The transition when DIR becomes as good as MER is depicted on the right of Figure 7.8. We see that the transition function is log-log linear with respect to d, where the slope depends on γ (the higher γ is, the more spacial reuse is possible, and transition occurs later). An example of the optimal schedule for the line case with $d = 3$ is given on the Figure 7.9.

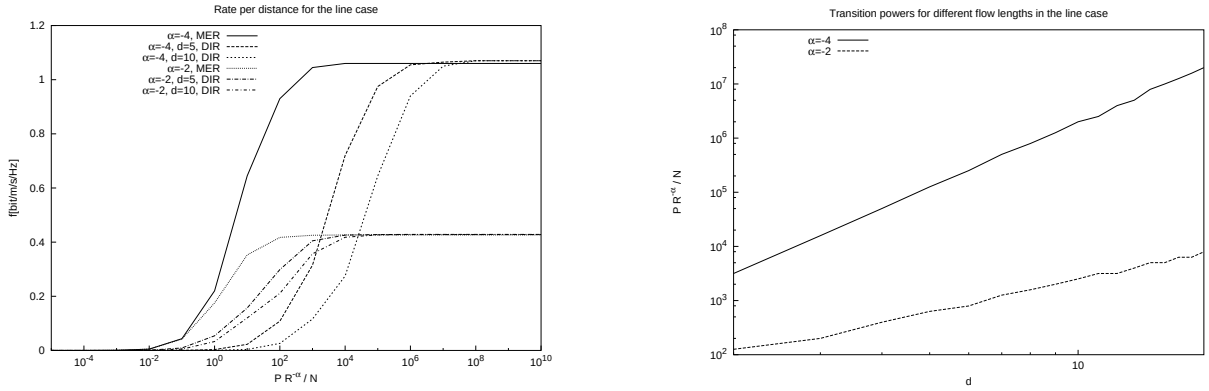


Figure 7.8: On the left, rates per distance achieved in the line case, for both routing policies, different flow lengths and power constraints. On the right, transition power for different flow length, on a log-log scale.

The fact that DIR is never better than MER in the line case is somewhat different from the ring case. In the ring case (Figure 7.4), each node was given a fixed time slot of a positive length. When transmitting power was high, optimal power allocation contained only one active node, and by increasing the transmission power it was able to increase arbitrary its rate. This was true for both MER and DIR, whereas the DIR rate in that case grew faster since relaying was expensive. In the line case, this scenario is no longer possible, since we always have more than one node active at the same time and the increase in the transmission powers will eventually be canceled by interference.

Similar phenomenon can be observed on a transition diagram for the line (Figure 7.8 on the right) and for the ring (Figure 7.5 on the right). For the line case, as the flow length increases, a transition power increases, whereas for the ring this dependency is inverse. This is again because for finite rings and DIR routing, the higher d is, the smaller the power for which optimal power

P^{MAX}	MER	s	DIR	(s,t)	Better
10^{-2}	**...1010...**	2	...111000111000...	(6,3)	MER
1	**...100100...**	3	...11100001110000...	(7,3)	MER
10	**...100100...**	3	...11000001100000...	(7,2)	MER
100	**...10001000...**	4	...1000000010000000...	(8,1)	MER
10^3	**...10001000...**	4	...100000000100000000...	(9,1)	MER
10^4	**...1000010000...**	5	...10000000000010000000000...	(11,1)	MER
10^5	**...1000010000...**	5	**...1000000000000100000000000...**	(13,1)	both
10^6	**...1000010000...**	5	**...1000000000000100000000000...**	(13,1)	both

Figure 7.9: Optimal scheduling in the line case, for different power limits, for both MER and DIR routing (d=3), in the same format as on the Figure 7.6. Dots before and after an array of ones and zeros means that the schedule consists of the same pattern infinitely repeating. While MER routing is better in the first six cases, both routing performs the same in the last two cases.

allocation has only one active node is, and the higher the derivative of the rate over transmitting power is (as can be seen on Figure 7.4). As it is not possible to have only one active node in an infinite network, the transition diagram changes and relaying becomes cheaper (a transition power grows with d).

Since in the line case MER is always optimal, we conclude that rate per distance is constant, regardless of the flow length. This in turn means that network transport capacity is roughly $\Theta(L)$ compared to the transport capacity $\Theta(\sqrt{L})$ of an arbitrary network on a unit disk, as in [30]. This is due to the fact that in our networks traffic is distributed locally ($d \ll L$) whereas in [30] sources and destinations are distributed uniformly throughout the network.

Finally, we analyze the optimal schedule for the line case. On Figure 7.10 we see how it evolves as L grows, for MER. For small powers we have $s = 2$. We see that as L grows, a power for which we have only one active nodes grows. For powers slightly lower than this transition one, we see that s is approximately 4.5. In the line case, we also have $s = 2$ for small powers, and it converges to around 4.5 as the power goes to infinity, as can be seen on the left of Figure 7.11. Since we allow only integer average distances in our approximate model (that is, all adjacent groups of active nodes are equally spaced), this limiting s is approximated with 5. We also see from Figure 7.11 that the same phenomenon occurs for $\gamma = 2$, hence it is not sensitive to the fading factor.

In the case of DIR routing, a similar behavior can be observed. For very small powers, the distance between groups of active nodes is $s = 2d$, $t = d$ and with this allocation each node either sends or receives data. As the power grows, s increases and d decreases. As the power tends to

infinity, t becomes 1 and s becomes $4.5 * d$. Therefore, for very large powers we can approximate a line (l, d) with a line $(ld, 1)$. Here we also see why there is a difference between rates on the left of Figure 7.8. Specifically, optimal s is 4.5 and in the MER case this is approximated with 5; it is thus obvious that in the DIR case, the error of the integer approximation of the optimal $s = 4.5 * d$ will be smaller than in the MER case, hence the rate will be higher. Again, the conclusions are the same for $\gamma = 2$.

We can again relate the optimal s to the size of the exclusion region. For simplicity of the presentation, let us consider only the MER case, since MER is always the optimal routing in the line case. Then, following the previous discussion, the size of the exclusion region is equal to $s - 1$. Hence we see that the size of the exclusion region increases with the transmission power, until it saturates to a fix value. This is much in contrast with the finding about the wide-band case, where the size of the exclusion region never saturates and always grows with power.

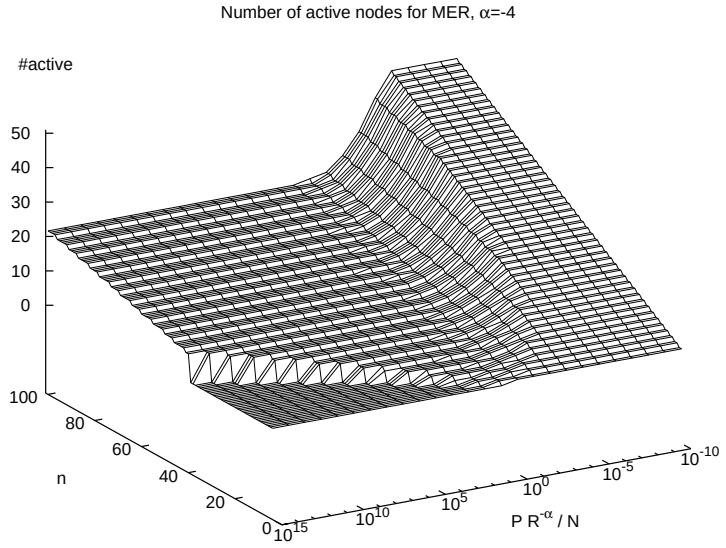


Figure 7.10: Number of nodes transmitting at the same time (inverse of s) for MER. For large n it converges to between 4 and 5 nodes. Since distance between active nodes grows with power, DIR will have the same scheduling behavior for large powers. Note that the flow length does not change scheduling policy for minimal energy routing, but only total rate.

In the line case, as in the ring case, we can analyze the energy dissipated by a node by counting a fraction of active nodes in the same slot. On the right of Figure 7.11 we see the fraction of nodes active in one slot. Since all active nodes sends with the same power and since this fraction is constant in all slots, it is directly proportional to the energy a node dissipates. Optimal power allocation for MER does not depend on flow length d, hence neither does energy

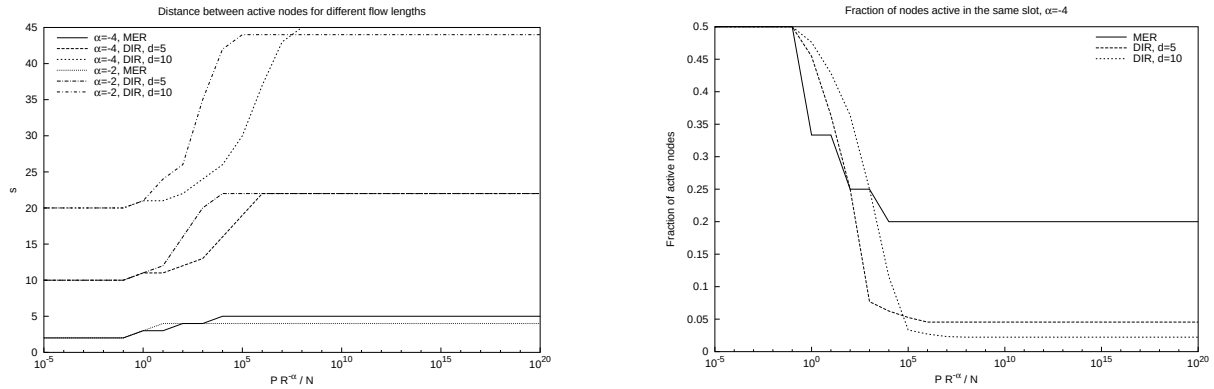


Figure 7.11: On the left is the average distance s between adjacent groups of active nodes. For small powers, s is 2 for MER and $2d$ for DIR. It converges to approximately 4.5 for MER and $4.5d$ for DIR. On the right is the fraction of nodes that are active in the same slot, which is directly proportional to dissipated energy.

dissipated. The total rate however decreases as d increases due to relaying. On the other hand, for high powers DIR can achieve the same rates as MER, but using d times less energy. For low powers DIR achieves lower rates than MER with approximately the same energy dissipation.

7.4 Networks in Plane

7.4.1 General Findings

The optimization problem (7.2) remains highly complex, and is difficult to solve in the general case, even using advanced optimization methods such as in [44]. A discussion for arbitrary networks with up to 6 nodes is given in [88]. It is impossible to draw general conclusions about network design from such small networks. Instead, we will use the findings from the linear case in order to derive heuristics that will find an approximative solution to the optimization problem in the general case.

In this section, we consider arbitrary networks in plane, as described in Section 7.2.3. We first consider power adaptation. As shown in Section 7.3, $0 - P^{MAX}$power control is optimal for linear networks. In other words, when a node sends data, it should always send with the full power. In Section 6 we theoretically proved that power control is not needed when the rate function is linear. However, the two findings do not apply directly to the case considered here. We further tested the optimality of $0 - P^{MAX}$power control numerically on a large number of random and hand-crafted examples with a small number of nodes and we found that it is always optimal or nearly optimal. We thus use this power allocation strategy in the rest of the analysis.

We next consider a random, two-dimensional network, with 40 nodes. The heuristics we use to solve the optimization problem is to define a number of alternative strategies for scheduling, and analyze the performance of each of them; we use the same heuristics as described in Section 5.6. Some of the strategies have variable exclusion region size, and others have fixed. We find that there should be an exclusion region around each destination receiving a packet. During packet reception, all nodes within this exclusion region should remain silent. Nodes outside of the exclusion region should transmit in parallel (to maximize spatial reuse). Also, it is easy to demonstrate that, within this framework, a sender should adapt the rate (that is, the modulation and coding) to the SINR experienced at the receiver.

The size of the optimal exclusion region does not depend on the size of a link, nor on the position of other nodes but only on the maximum transmitting power. At first glance, this seems in contradiction with the results from Section 7.3, where we showed that the size of the exclusion region in case of narrow-band networks does depend on sizes of links. However, in the scenarios considered, sizes of links do not vary significantly, hence we cannot observe the sensitivity of the exclusion region to these variations.

7.4.2 Application to 802.11

For the physical layer described in [3] and channel described in [28], we found that the optimal strategy is to exclude link k, when link l is receiving, if $P_k^{MAX} h_{kl}/\eta > 17dB$, regardless of link sizes, positions of other nodes and transmitted power. This is in sharp contrast to the existing rate adaptation protocols [35, 33, 43, 50] that send RTS/CTS packets with the lowest available rate (around 0dB), hence they induce larger, suboptimal exclusion regions.

In order to illustrate this result, we give in detail the comparison of the optimal strategy to three other common strategies. One strategy is to let all nodes transmit all the time (interference is always allowed). The other is to have only one node at a time (this is, in our framework, equivalent to time division multiple access, TDMA). Finally, we emulate the 802.11 MAC. Standard RTS/CTS packet in 802.11 are sent with the lowest possible code and modulation rate so it can be decoded by a large number of nodes. The actual threshold varies with the deployed hardware and we set the required SNR of the RTS/CTS to 0dB. Before every transmission, a source and a destination will send RTS/CTS. All the nodes that receive RTS/CTS with SNR higher than 0dB will decode it; they will be informed about the incoming transmission, and will remain silent. The others nodes will transmit in parallel. We do not simulate RTS/CTS packet but we require inequality $P_k^{MAX} h_{kl}/\eta < 0dB$ to be satisfied for concurrently transmitting nodes.

Given a network and a traffic demand, we first choose the scheduling strategy. We then fix

the maximum hop length for the routing algorithm, and obtain the routing matrix R. For each link we construct a list of links that cannot be active at the same time (due to exclusions). We repeat the same procedure for each node and each flow, and construct a schedule in a greedy manner. Then, we optimize slot frequencies in order to maximize the system utility. We repeat the same for each routing matrix and we choose the routing that maximizes the performance.

We numerically analyze the performance of these strategies on a set of networks with 40 nodes, where nodes are randomly distributed on 200m x 200m square. We take the [28] NLOS path loss model. We consider uniform and non-uniform network topologies, both in the sense of node positioning, power constraints and traffic demands. The results of the numerical comparisons are presented on Figure 7.12.

We first consider a network of uniformly distributed nodes on a unit square, where half of the nodes are sources talking to a randomly chosen destinations from the other half of the nodes. The results are given on the top of Figure 7.12. We see that in most of the cases the 17 dB strategy is the optimal one. The only exception is in case of 1mW transmitted power where the optimal exclusion region should be even slightly smaller. We see that the utility of the optimal strategy is by 5 to 10 higher than the utility of 802.11 exclusion region. For a network with 40 nodes (20 flows) this means that we have approximately from 30% to 60% of rate improvement per flow. We also consider heterogeneous scenarios in the middle and on the bottom of Figure 7.12, and we see that the same conclusions hold in those cases as well.

In conclusion, we find that the optimal MAC is similar to the existing 802.11 RTS/CTS protocol. However, these protocols do not control the size of the exclusion region. They send RTS/CTS packets with the lowest possible rate, thus making the exclusion regions too large, and reduce the spatial reuse. We find that, for the rate function of [3] and channel model of [28], the optimal RTS/CTS packets should be coded for SNR at the receiver of 17 dB instead of 0dB. Similar analysis can be performed for different rate functions. In, general, the existing 802.11 could be improved by reducing the exclusion region, i.e., allowing more concurrent transmissions (and thus more interference).

The conclusions presented in this section hold both for DIR and MER routing. We do not compare the performances of DIR and MER routing here. DIR routing is very simple, whereas MER routing assumes relaying and an additional protocol complexity that we do not model here. Therefore, we leave more detailed comparison of DIR and MER routing on arbitrary networks in plane for a future work.

7.4.3 Other Examples

We applied our model to the cellular setting by letting all receivers concentrate in one point, representing a base station. We obtained the same conclusion as [48]: near-by nodes should send separately (they are in the exclusion region of the base-station so they have to time-share), while distant nodes should send together (since they are outside of the exclusion region). Rates are also adapted to the channel condition and the transmitted power is always maximum. Our findings thus confirm the results from [48].

Finally, we also compared the optimal MAC with CA/CDMA [62], the state-of-art power adaptation protocol for fixed rate networks. We implement a simplified version that does not account for protocol overhead but only reflects the fundamental principles of the protocol. We found that CA/CDMA performs far worse than the optimal version; it cannot benefit from power adaptation. We omit the numerical results for these two cases since the findings are similar to the 802.11 case.

7.5 Summary of Main Findings and Conclusions

We have given a general model for a joint scheduling, power allocation and routing optimization problem. We solved the problem for symmetric one-dimensional networks, for both direct and minimum energy routing policies. Using the findings from one-dimensional networks, we propose heuristics for optimizing performance of arbitrary networks in plane. Our main findings are

Optimal Power Control is $0 - P^{MAX}$ - Whenever a node is transmitting, it should transmit with the maximum power. This is indeed what is implemented in the 802.11 MAC. However, this results is in contrast to a class of a power optimization protocols (e.g. [61]). We show that when a rate can be adapted to SINR at a receiver, it is optimal to adapt rates and keep transmission powers fixed.

Optimal Scheduling is Based on the Exclusion Regions - We characterized the optimal scheduling and power allocation for both routing policies. We find that, like in the wide-band case, the optimal schedule is based on the exclusion regions: while a node receives, there should be no transmitter active in the vicinity of the receiver.

Optimal Size of Exclusion Regions - The size of the exclusion region depends on the transmission power and, unlike in the wide-band case, on link lengths. For large networks and large

power constraints, group sizes converge to 1 and the distance between groups converges to 4.5 times the size of a used link (1 for MER, d for DIR).

It is Better to Relay With Small Transmission Powers - We found that for small power constraints it is better to relay, and for large power constraints it is better to send data directly to destinations. We also describe for which power constraints this transition occurs. In the case of an infinitely large network, minimum energy routing is always as good as direct routing. However, in contrast with their names, direct routing dissipates less energy than minimum energy routing for large powers, while achieving the same rate. This finding is in contrast with the wide-band case, where MER routing is always optimal, regardless of power constraints.

The Optimal MAC Depends on the Choice of Routing - Since the size of the exclusion region depends on the length of a link, which in turn depends on the choice of routing, it follows that the optimal choice of MAC is dependant on the choice of routing, unlike in the wide-band case.

802.11 RTS/CTS is Suboptimal - The actual size of the exclusion regions, as implemented in 802.11 MAC, is suboptimal. The exclusion regions are too large, and the spatial reuse is too low. We find that it is optimal to exclude only nodes that receive RTS/CTS signals with an SNR higher than 17 dB, which significantly decreases the size of exclusion regions.

We do not solve exactly the proposed optimization problem. Our results are heuristics based on numerical explorations. Nonetheless, even in this current form, they do show interesting conclusions for improving existing MAC protocols. We hope our findings will trigger further analysis to provide theoretical confirmation, as is already available in some special cases [19, 48].

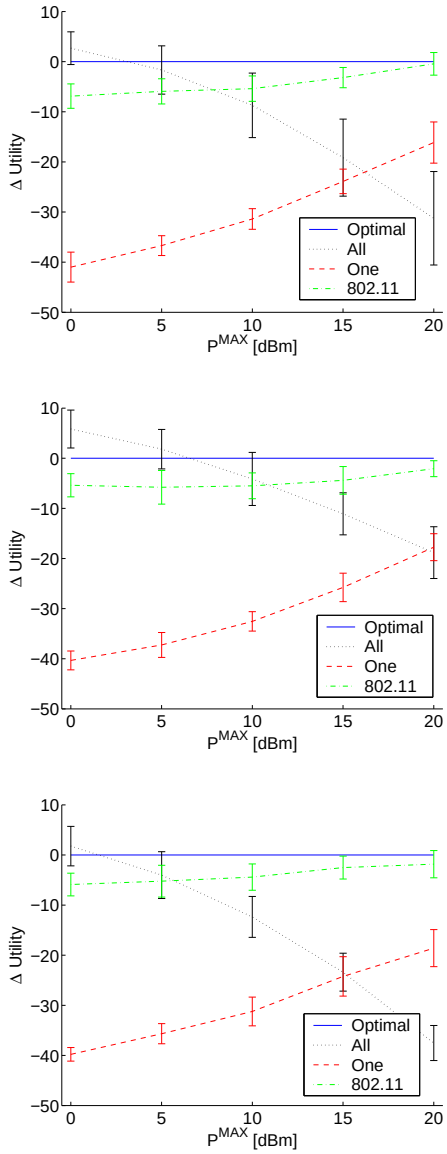


Figure 7.12: Performance comparison of different MAC protocols on random networks with 40 nodes on 200m x 200m square. On the x axis: we plot the maximum transmitted power constraint. On the y axis we plot the difference between the log-utility achieved by the optimal protocol and those achieved by different other protocols. On the top, we consider topologies where nodes are distributed uniformly random on the square, and sources and destinations of flows are randomly chosen among them. In the center, we assume 1/4 of nodes is distributed on the left half of the square and 3/4 on the right half. Sources and destinations of flows are again randomly chosen among them. Finally, in the bottom, nodes, sources and destinations are uniformly distributed, but the power constraints of nodes are not uniform and vary randomly between 50% and 150% of the constraint given on the x axis.

Chapter 8

Conclusions

Conclusions

In this thesis, we consider a cross-layer design of multi-hop wireless ad-hoc networks. We analyze the jointly optimal design of the medium access, physical layer and routing.

We map a network design into an optimization problem. The main building blocks of a multi-hop wireless network are routing, medium access, power control and rate control. We present a mathematical model where each building block is modeled as a set of free variables. Possible values of these variables are constrained by different design requirements, hardware and regulatory constraints. Using this mathematical model, we characterize a set of end-to-end flow rates that are feasible under these constraints. The optimal network design is then a solution to an optimization problem, which aims at maximizing a certain design objective.

In the first part of the thesis, we consider different possible design objectives. We focus on rate-based design objectives. The four most frequently used design objectives for wireless networks are maximizing total capacity, maximizing transport capacity, max-min fairness and utility fairness.

We first analyze max-min fairness. We observe that in many examples arising in the context of wireless networks, it is not clear whether max-min fair rate allocation exists, and if it exists, whether it can be obtained using existing algorithms, such as water-filling. We propose a general framework for max-min fairness on Euclidean spaces, and we define a class of sets in which we show max-min fairness always exist. We also define a centralized algorithm that finds the max-min fair point whenever it exists. We then analyze the feasible rate sets of the examples for which the existence of max-min fairness has not been previously shown. In all these examples we are able to show, using our results, that the max-min fair allocation exists, and we give an

algorithm that locates it.

Once we have shown that the max-min fair allocation exists in the wireless scenarios of our interest, we continue by analyzing properties of the four rate-based design objectives. We show that max-min fairness is a very inefficient objective. In all cases it leads to equal rates of all flows, the same as the rate of the worst flow. We also show that maximizing the sum of rates is an unfair metric: all flows but the best will have zero rate. The same problem exists with transport capacity, or any metric that maximizes a weighted sum of rates. We finally analyze several utility-based metrics. We show that one of them, proportional fairness, exhibits a good trade-off between efficiency and fairness, and it is an ideal candidate for a design objective for multi-hop wireless networks.

In the second part of the thesis we focus on the above defined optimization problem, taking proportional fairness as the optimization criteria. We first analyze the optimization problem for the case of wide-band, and then for narrow-band networks.

In the case of wide-band networks, it can be shown that a rate of a link is a linear function of signal-to-interference-and-noise ratio at the receiver. We use this specificity of the wide-band physical layer to solve the optimization problem. We find that the optimal solution is characterized by the following rules: (1) When data is being sent over a link, it is optimal to have an exclusion region around the destination, in which all nodes remain silent during transmission, whereas nodes outside of this region can transmit in parallel, regardless of the interference they produce at the destination. Additionally, the source adapts its transmission rate according to the level of interference at the destination due to sources outside of the exclusion region. (2) The optimal size of this exclusion region depends only on the transmission power of the source of the link, and not on the length of the link nor on the positions of the nodes in its vicinity. (3) Each node in a given time slot either sends data at the maximum power, or does not send at all. As for the routing, we restrict ourselves to a subset of routes where on each successive hop we decrease the distance toward the destination, and we show that (4) relaying along a minimum energy and loss route is always better than using longer hops or sending directly. Finally (5), the design of the optimal MAC protocol is independent of the choice of the routing protocol.

The third finding we prove analytically for an arbitrary network. We further generalize it for an arbitrary rate-based design objective, and also show that any other power control strategy that is not of this form, is strictly suboptimal, regardless of what design objective we use. Finally, we show that this power control strategy is optimal for energy-minimization objectives as well.

The rest of the findings for the wide-band case we show numerically. We first show the optimization problem exactly for 1-D cases. Using these results, we then define heuristics that

approximatively solve the optimization problem for an arbitrary network in plane. In all cases we confirm our findings (1)-(5).

One of the interesting consequences of our findings is an application to low-power ultra-wide band networks. From (2) it follows that when the transmitting power is sufficiently low, the exclusion region diminishes, and there is no need for complex mutual exclusion protocols. We verify this finding on a realistic UWB physical layer. We propose a simple interference mitigation technique that mitigates the interference from near-by interferers. We then show that for a PPM-UWB physical layer, augmented with the interference mitigation techniques, it indeed is optimal not to enforce any exclusions but, instead, to let all links transmit concurrently whenever they have something to send.

Finally, we discuss narrow-band networks. We find that in the narrow-band case, findings (2), (4) and (5) do not hold. Instead, we find that the choice of routing depends on the transmitting power. When the transmitting power is large, it is better to send directly to destinations, whereas for small powers it is better to relay over minimum energy routes. We again find that it is optimal to have exclusion regions around destinations. However, the optimal sizes of the exclusion regions depend not only on transmitting power, but also on link sizes. Consequently, they will depend on the choice of routing, hence in narrow-band networks it is not always possible to decouple routing and MAC, as in the wide-band case.

The findings from narrow-band case are not exact and are obtained using different heuristics. However, they can be used to improve the performance of the existing narrow-band protocols. Using our findings we show that the exclusion mechanism in 802.11 is not optimal. It strives to eliminate most of the interference, hence the exclusion regions are large, which yields small spatial reuse. Instead, one should decrease the sizes of the exclusion regions, allow more interference, and adapt rates to the interferences at the receivers. We give an approximatively optimal size of the exclusion regions for 802.11 case and numerically show that this leads to large improvements in rates.

Future Work

Most of the research in this thesis is focused on solving the optimization problem on theoretical models of networks. These models however assume that a network is able to provide a perfect exchange of control information among nodes, which is rarely the case. In the analysis presented in the thesis, this assumption is justified by the fact that we want to find the optimal design guidelines for a network, and not to model the inefficiencies of existing network protocols.

The next step in the research is to analyze the effect of our design decisions on protocol implementations. The first step in this direction we have taken in Section 5.8, where we analyze low-power UWB networks; also in Section 7.4, where we propose modifications to the 802.11 MAC.

It is then not clear how such an exclusion and rate adaptation protocol would be implemented in practice, and if the cost of such a protocol would overcome its benefits. This is especially interesting for the cases where exclusions are needed.

It also remains to evaluate in more depth a relation between routing protocols and PHY/MAC layers in a realistic setting, again considering the implementation cost of these protocols.

Finally, all of these protocols are analyzed with rate-based design objectives in mind. It remains as a future work to analyze design objectives whose goal is to minimize energy dissipation, and to see how these types of design objectives would influence the optimal cross-layer design of a wireless multi-hop network.

Bibliography

[1] IEEE 802.15.4a approved technical requirements document, may 04 - 04/198r2, available at http://www.ieee.org.

[2] IEEE 802.15 WPAN High Rate Alternative PHY Task Group 3a. Multi-band ofdm physical layer proposal. *IEEE 802.15-03/267r5*, July 2003.

[3] http://www.connectronics.com/tsunami/mp11a_0803.pdf.

[4] P. Baldi, L. De Nardis, and M.-G. Di Benedetto. Modeling and optimization of uwb communication networks through a flexible cost function. *IEEE Journal on Selected Areas in Communications*, 20(9):1733–44, December 2002.

[5] N. Bambos, S. Chen, and G. Pottie. Radio link admission algorithms for wireless networks with power control and active link quality protection. In *Proceedings INFOCOM*, 1995.

[6] A. et al. Bedekar. Downlink scheduling in cdma data networks. In *Globecom'99*, 1999.

[7] G. Berger-Sabbatel, F. Rousseau, M. Heusse, and A. Duda. Performance anomaly of 802.11b. In *INFOCOM'03*, San Francisco, California, April 2003.

[8] D. Bertsekas and R. Gallager. *Data Networks*. Prentice-Hall, 1987.

[9] S. Borst and P.A. Whiting. Dynamic rate control algorithms for hdr throughput optimization. In *INFOCOM*, 2001.

[10] W. Bossert and J.A. Weymark. Utility in social choice. In S. Barbera, P.J. Hammond, and C. Seidl, editors, *Handbook of Utility Theory*. Kluwer Academic Publishers, 2004.

[11] P. Brémaud. *Markov chains, Gibbs fields, Monte Carlo simulation, and Queues*. Springer-Verlag, 2001.

[12] M. Cagalj, J.-P. Hubaux, and C. Enz. Energy-efficient broadcasting in all-wireless networks. *To appear in ACM/Kluwer Mobile Networks and Applications (MONET)*, 2004.

[13] Z. Cao and E. Zegura. Utility max-min: An application-oriented bandwidth allocation scheme. In *INFOCOM'99*, pages 793–801, 1999.

[14] J.H. Chang and L. Tassiulas. Energy conserving routing in wireless ad-hoc networks. In *INFOCOM'00*, pages 22–31, 2000.

[15] A. Charny. An algorithm for rate allocation in a packet-switched network with feedback. *M.S. thesis*, MIT, May 1994.

[16] M.A. Chen. Individual monotonicity and the leximin solution. *Economic Theory*, 15:353–365, 2000.

[17] ATM Forum Technical Committee. Traffic management specification - version 4.0. *ATM Forum/95-0013R13*, February 1996.

[18] T. Cover and J.A. Thomas. *Elements of Information Theory*. John Whiley & Sons, 1991.

[19] R. Cruz and A.V. Santhanam. Optimal link scheduling and power control in cdma multihop wireless networks. In *Globecom'02*, 2002.

[20] R. Cruz and A.V. Santhanam. Optimal routing, link scheduling and power control in multihop wireless networks. In *Proceedings INFOCOM'03*, 2003.

[21] F. Cuomo, C. Martello, A. Baiocchi, and F. Capriotti. Radio resource sharing for ad hoc networking with uwb. *IEEE Journal on Selected Areas in Communications*, 20(9):1722–1732, December 2002.

[22] G. Durisi and G. Romano. On the validity of gaussian approximation to characterize the multiuser capacity of UWB TH PPM. In *IEEE Conference on Ultra Wideband Systems and Technologies*, pages 151–161, 2002.

[23] E. Hahne. "Round-Robin Scheduling for Max-Min Fairness in Data Networks". *IEEE Journal on Selected Areas in Communications*, 9(7):1024–1039, Sept. 1991.

[24] T. Elbatt and A. Ephremides. Joint scheduling and power control for wireless ad-hoc networks. In *Proceedings INFOCOM*, 2002.

[25] G. J. Foschini and Z. Miljanic. A simple distributed autonomous power control algorithm and its convergence. *IEEE Transactions on Vehicular Technology*, 42(4):641–646, November 1993.

[26] P.K. Frenger, P. Orten, T. Ottosson, and A.B. Svensson. Rate-compatible convolutional codes for multirate DS-CDMA systems. *IEEE Transactions on Communications*, 47(6):828–836, June 1999.

[27] V. Gambiroza, B. Sadeghi, and E. Knightly. End-to-end performance and fairness in multihop wireless backhaul networks. In *Technical Report*, 2004.

[28] S.S. Ghassemzadeh and V. Tarokh. Uwb path loss characterization in residential environments. In *IEEE Radio Frequency Integrated Circuits (RFIC) Symposium*, pages 501–504, June 2003.

[29] A. Goldsmith and S.B. Wicker. Design challenges for energy-constrained ad hoc wireless networks. *IEEE Wireless Communications*, 9(4):8–27, August 2002.

[30] P. Gupta and P.R. Kumar. The capacity of wireless networks. *IEEE Transactions on Information Theory*, 46(2):388–404, March 2000.

[31] H. Tzeng and K. Siu. "On Max-Min Fair Congestion Control for Multicast ABR Service in ATM". *IEEE Journal on Selected Areas in Communications*, 15(3):545–556, April 1997.

[32] J. Hagenauer. Rate-compatible punctured convolutional codes (RCPC codes) and their applications. *IEEE Transactions on Communications*, 36(4):389–400, April 1988.

[33] I. Haratcherev, K. Langendoen, R. Lagendijk, and H. Sips. Hybrid rate control for IEEE 802.11. In *MobiWac*, 2004.

[34] W. Hirt and D. Porcino. Pervasive Ultra-wideband Low Spectral Energy Radio Systems (PULSERS), White Paper. *Integrated Project PULSERS, http://www.pulsers.net*, November 2002.

[35] G. Holland, N. Vaidya, and P. Bahl. A rate-adaptive max protocol for multi-hop wireless networks. In *Proceedings MobiCom'01*, Rome, Italy, July 2001.

[36] Hou, Y. and Tzeng, H. and Panwar, S. A generalized max-min rate allocation policy and its distributed implementation using the ABR flow control mechanism. In *INFOCOM'98*, pages 1366–1375, 1998.

[37] X. Huang and B. Bensaou. On max-min fairness and scheduling in wireless ad-hoc networks: Analytical framework and implementation. In *Proceedings MobiHoc'01*, Long Beach, California, October 2001.

[38] IEEE P802.15 Working group. The uwb indoor path loss model. *Technical report P802.15-02/277r0-SG3a*, June 2002.

[39] J. Byers, et al. "A Digital Fountain Approach to Reliable Distribution of Bulk Data". In *ACM SIGCOMM '98*, September 2-4 1998.

[40] J. Ros and W. Tsai. "A Theory of Convergence Order of Maxmin Rate Allocation and an Optimal Protocol". In *INFOCOM'01*, pages 717–726, 2001.

[41] R. Jain. *The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation, and Modeling*. Wiley-Interscience, 1991.

[42] D.B. Johnson and D.A. Maltz. Dynamic source routing in ad hoc wireless networks. In T. Imielinski and H. Korth, editors, *Mobile Computing*, chapter 5, pages 153–181. Kluwer Publishing Company, 1996.

[43] H.-J. Ju, I. Rubin, and Kuan Y-C. An adaptive RTS-CTS control mechanism for IEEE 802.11 MAC protocol. In *VTC*, 2003.

[44] D. Julian, M. Chiang, D. O'Neill, and S. Boyd. Qos and fairness constrained convex optimization of resource allocation for wireless cellular and ad hoc network. In *INFOCOM*, 2002.

[45] E.-S. Jung and N. H. Vaidya. An energy efficient mac protocol for wireless lans. In *Proceedings INFOCOM*, 2002.

[46] V. Kawadia, S. Narayanaswamy, R. Rozovsky, R.S. Sreenivas, and P.R. Kumar. Protocols for media access control and power control in wireless networks. In *Proceedings of the 40th IEEE Conference on Decision and Control*, pages 1935–1940, Orlando, FL, Dec. 4–7 2001.

[47] F. P. Kelly, A.K. Maulloo, and D.K.H. Tan. Rate control in communication networks: shadow prices, proportional fairness and stability. *Journal of the Operational Research Society*, 49:237–252, 1998.

[48] K. Kumaran and L Qian. Uplink scheduling in CDMA packet-data systems. In *INFOCOM*, 2003.

[49] L. Georgadis, et al. Lexicographically optimal balanced networks. In *INFOCOM'01*, pages 689–698, 2001.

[50] M. Lacage, M. Manshaei, and T. Turletti. IEEE 802.11 rate adaptation - a practical approach. In *MSWiM*, 2004.

[51] J.-Y. Le Boudec. A tutorial on rate adaptation, congestion control and fairness in the internet. *Technical Report, http://ica1www.epfl.ch/PS_files/LEB3132.pdf*, 2000.

[52] X. Lin and N. Shroff. Joint rate control and scheduling in multihop wireless networks. In *43rd IEEE Conference on Decision and Control*, Bahamas, December 2004.

[53] S. Low and D. Lapsley. Optimization flow control, I: basic algorithm and convergence. *IEEE/ACM Transactions on Networking*, December 1999.

[54] A. Mas-Colell, M. Whinston, and J. Green. *Microeconomic Theory*. Oxford University Press, 1995.

[55] L. Massoulie and J. Roberts. Bandwidth sharing: objectives and algorithms. *IEEE/ACM Transactions on Networking*, 10(3):320–328, June 2002.

[56] Related Matlab scripts, http://icapeople.epfl.ch/bradunov/jsac.zip.

[57] A.L. McKellips and S. Verdu. Maximin performance of binary-input channels with uncertain noise distributions. *IEEE Transactions on Information Theory*, 44(3):947–972, May 1998.

[58] R. Merz, J. Widmer, J.-Y. Le Boudec, and B. Radunovic. A joint PHY/MAC architecture for low-radiated power TH-UWB wireless ad-hoc networks. *Wireless Communications and Mobile Computing Journal, Special Issue on Ultrawideband (UWB) Communications*, June 2005.

[59] J. Mo and J. Walrand. Fair end-to-end window-based congestion control. *IEEE/ACM Transactions on Networking*, 8(5):556–567, Oct. 2000.

[60] J. P. Monks, V. Bharghavan, and W.W. Hwu. A power controlled multiple access protocol for wireless packet networks. In *Proceedings INFOCOM*, 2001.

[61] A. Muqattash and M. Krunz. CDMA–based MAC protocol for wireless ad hoc networks. In *Proceedings of MOBIHOC'03*, pages 153–164, June 2003.

[62] A. Muqattash and M. Krunz. A single-channel solution for transmission power control in wireless ad hoc networks. In *Proceedings of MOBIHOC'04*, May 2004.

[63] T. Nandagopal, T.E. Kim, X. Gao, and V. Bharghavan. Achieving MAC layer fairness in wireless packet network. In *MOBICOM*, 2000.

[64] S. Narayanaswamy, V. Kawadia, R. S. Sreenivas, and P.R. Kumar. Power control in ad-hoc networks: Theory, architecture, algorithm and implementation of the COMPOW protocol. In *European Wireless 2002*, pages 156–162, February 2002.

[65] M. Neely, E. Modiano, and C. Rohrs. Dynamic power allocation and routing for time varying wireless networks. *IEEE Journal on Selected Areas in Communications*, 23(1):89–103, January 2005.

[66] M.J. Neely, E. Modiano, and C.E. Rohrs. Dynamic power allocation and routing for time varying wireless networks. In *INFOCOM*, 2003.

[67] P. Bender, et al. CDMA/HDR: A bandwidth-efficient high-speed wireless data service for nomadic users. *IEEE Communications Magazine*, pages 70–77, July 2000.

[68] P. Marbach. "Priority Service and Max-Min Fairness". In *INFOCOM'02*, 2002.

[69] C.E. Perkins and E.M. Royer. Ad-hoc on-demand distance vector routing. In *In Proceedings of IEEE WMCSA'99*, pages 90–100, New Orleans, LA, February 1999.

[70] X. Qiu and K. Chawla. On the performance of adaptive modulation in cellular systems. *IEEE Transactions on Communications*, 47(6):884–895, June 1999.

[71] R. Horst, P. Pardalos, N. Thoai. *Introduction to Global Optimization*. Kluwer Academic Publishers, 2000.

[72] B. Radunovic. A cross-layer system simulator for UWB-based wireless sensor networks. *IBM Research Report RZ 3594*, February 2005.

[73] B. Radunović and J. Y. Le Boudec. Optimal power control, scheduling and routing in UWB networks. *IEEE Journal on Selected Areas in Communications*, September 2004.

[74] B. Radunovic, L. Truong, and M Weisenhorn. On architectures of transmit-only, UWB-based wireless sensor networks. *IBM Research Report RZ 3596*, 2005.

[75] A. Rajeswaran and R. Negi. Capacity of power constrained ad-hoc networks. In *Proceedings INFOCOM*, 2004.

[76] J. Rawls. *A Theory of Justice*. Harvard University Press, 1971.

[77] V. Rodoplu and T.H. Meng. Minimum energy mobile wireless networks. *IEEE J. Select. Areas Commun.*, 17(8):1333 – 1344, August 1999.

[78] D. Rubenstein, J. Kurose, and D. Towsley. The impact of multicast layering on network fairness. *IEEE/ACM Transactions on Networking*, 10(2):169–182, Apr. 2002.

[79] D. Rus. Keynote on autonomous mobile networks. In *The First IEEE Workshop on Embedded Networked Sensors (EmNetS-I)*, 2004.

[80] S. Boyd, L. Vandenberghe. *Convex Optimization*. http://www.stanford.edu/class/ee364/, 2003.

[81] B. Sadeghi, V. Kanodia, A. Sabharwal, and E. Knightly. Opportunistic media access for multirate ad hoc networks. In *MOBICOM*, September 2002.

[82] S. Sarkar and L. Tassiulas. Fair allocation of discrete bandwidth layers in multicast networks. In *INFOCOM'00*, pages 1491–1500, 2000.

[83] Y. Souilmi, R. Knopp, and G. Caire. Coding strategies for UWB interference- limited peer-to-peer networks. In *WiOpt'03: Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks*, INRIA Sophia-Antipolis, France, March 3-5 2003.

[84] A. Tang, J. Wang, and S. Low. Is fair allocation always efficient. In *INFOCOM*, 2004.

[85] L. Tassiulas and A. Ephremides. Jointly optimal routing and scheduling in packet radio networks. *IEEE Transactions on Information Theory*, 38(1):165–168, January 1992.

[86] L. Tassiulas and S. Sarkar. Max-min fair scheduling in wireless networks. In *Proceedings INFOCOM'02*, 2002.

[87] I.E. Telatar and D.N.C. Tse. Capacity and mutual information of wideband multipath fading channels. *IEEE Transactions on Information Theory*, 46(4):1384–1400, 2000.

[88] S. Toumpis and A.J. Goldsmith. Capacity regions for wireless ad hoc networks. *IEEE Transactions on Wireless Communications*, 2(4):736–748, July 2003.

[89] S. Toumpis and A.J. Goldsmith. Performance, optimization, and cross-layer design of media access protocols for wireless ad hoc networks. In *International Conference on Communications (ICC)*, pages 2234–2240, Anchorage, AK, May 2003.

[90] D. Tse and S. Hanly. Multi-access fading channels - part i - polymatroid structure, optimal resource allocation and throughput capacities. *IEEE Transactions on Information Theory*, 44(7):2796–2815, November 1998.

[91] B. van der Wal et al. Definition of UWB scenarios, Deliverable D2a2 - Initial. *Integrated Project PULSERS, http://www.pulsers.net*, March 2004.

[92] J. Van Leeuwen. Graph algorithms. In J. Van Leeuwen, editor, *Algorithms and Complexity*. Elsevier, 1992.

[93] S. Verdu. Spectral efficiency in the wideband regime. *IEEE Transactions on Information Theory*, 48(6):1319–1343, 2002.

[94] M. Weisenhorn. Physical layer for reader scenario. *IBM Research Report RZ 3595*, 2005.

[95] M. Weisenhorn and W. Hirt. Robust noncoherent receiver exploiting UWB channel properties. In *IEEE Joint UWBST & IWUWBS 2004*, pages 156–160, Kyoto, May 2004.

[96] M. Win and R. Scholtz. Ultra-wide bandwidth time-hopping spread-spectrum impulse radio for wireless multiple-access communications. *IEEE Transactions on Communications*, 48(4):679–691, April 2000.

[97] R. D. Yates. A framework for uplink power control in cellular radio systems. *IEEE Journal on Selected Areas in Communications*, 13(7):1341–1347, September 1995.

[98] Y. Yi and S. Shakkottai. Dynamic power allocation and routing for time varying wireless networks. In *INFOCOM*, 2004.

Appendix A

Receiver Architectures for Transmit-Only, UWB-Based Sensor Networks

A.1 Introduction

The work presented in this chapter was done during an internship I did in IBM Research Labs in Zurich. We considered design of wireless sensor networks. Due to specific requirements of sensor networks, the imposed design constraints are more stringent. Therefore, this research is not directly related to the rest of the results presented in this thesis. However, since this thesis is on wireless network design, we include the results as an appendix.

There is a recent increase in interest for wireless sensor networks, because of their easy deployment and the resulting low cost. Those networks can serve for different purposes, from measurement and detection, to automation and process control. One such scenario, called sensor positioning and identification network (SPIN) is described in [34].

We consider sensor networks that are composed of three types of devices. The first type are *Sensor Nodes* (SNs) that sense the environment and transmit the resulting data. They may be deployed in high numbers hence they need to be very cheap and power efficient. SNs may be distributed over a large area, so intermediate devices are needed to relay the data packets. We call these devices *Cluster Heads* (CHs), as they are responsible for clustering the SNs. CHs receive the data packets sent by the SNs and transfer them to one or more *Central Servers* for further processing.

One of the main design considerations is low cost. To achieve this goal we choose UWB impulse-radio for the physical layer. Transmitters of pulsed UWB signals generate short pulses at certain points in time, therefore they are easy to implement and cheap. However, even simple

receivers for pulsed UWB signals are much more complex than the corresponding transmitter, the reason are the required complex functions like acquisition and synchronization. Therefor, a receiver may be prohibitively expensive for low-cost SNs; to enable low cost SNs we focus on transmit-only SNs. The concept of transmit-only sensor networks has been originally proposed within the PULSERS project [34].

Since SNs cannot sense the medium, they cannot prevent parallel transmissions that may lead to reception failures due interferences, this effect decreases the overall performance. We propose two different architectures for CHRs that address this problem:

The so called *adaptive* CHR is based on an adaptive threshold of the signal strength of the received packets. A CHR starts receiving a packet only if the packet's signal strength exceeds the threshold; the threshold is adapted to the network traffic which is constantly tracked by the CHR.

The second architecture assumes that each CH contains one receiving circuit and an additional acquisition circuit. A CH starts receiving the first packet it detects. The function of the additional acquisition circuit is to monitor the wireless medium and to detect if a packet, with higher signal strength appears than the one currently being received. If this happens, the receiving circuit drops the weaker packet and switches to the stronger one. We call this the *switched* CHR.

We analyzed the performance of the two proposed CHR architectures using a custom-designed simulator. We compared them with a *simple* CHR *architecture* having no adaptive threshold or switching mechanism. We find that, during traffic bursts, the adaptive-threshold approach yields great improvement compared to the non-adaptive one, while maintaining the same level of complexity. The switched architecture introduces an additional performance improvement but with a slight increase in CH implementation cost.

In Section A.2 we give a detailed description of the system model and performance objectives. Detailed descriptions of the proposed CHR architectures are given in Section A.3, and performance results are given in Section A.4. A conclusion given is in Section A.5. For a discussion of SN and CH architectures that goes beyond the subject of this paper see [74].

A.2 System Model and Performance Objectives

A.2.1 Sensor Node Architecture

We assume a network of transmit-only SNs, equipped with sensing and transmitting devices. They are designed for low data rates. Typical SN data rates are from 1 kbps up to 100 kbps (in case of audio or video transmissions). As mentioned abouve, SNs cannot sense the wireless medium nor can they receive any feedback from the CHs or other SNs. It may happen that packets from two or more SNs overlap and collide, while the SNs will have no feedback on the collision. To provide a uniform collision probability, SNs use an ALOHA-like random medium access scheme [74].

Packets are removed by the SNs right after their transmission, whether they were successfully received or not. Packet losses are resolved at the application layer, e.g. by having the sampled packets sent to the central servers and processed there [74]. This choice of SN architecture implies that most of the complexity is in the CHs.

A.2.2 Ultra Wideband Physical Layer

The considered sensor network architecture imposes three basic requirements to the physical layer (PHY). The aggregate traffic rate may be in the order of several megabits per second, since the traffic rate of a single sensor goes up to 100 kbps and as each CH has to support a large number of SNs. Secondly, we need a low-cost and simple PHY, as it is going to be implemented in a large number of SNs. Finally, another frequent requirement is location-awareness; various sensor applications need ranging and positioning capabilities [91].

In order to address the above requirements a non-coherent, pulse-based UWB physical layer with binary pulse position modulation (2PPM) is chosen. For a typical channel delay spread of 100 ns this modulation scheme allows to achieve a raw data rate of 5 Mbit/s. As mentioned a corresponding transmitter is relatively simple; low complexity of the CHs is achieved by the use of noncoherent receivers. Furthermore, the use of pulse-based UWB signals can support accurate ranging capabilities. More details about this UWB physical layer can be found in [94].

A.2.3 Cluster Head Design Objectives

We divide SNs into two classes: (i) high priority SNs generating important data and at high rates. (ii) Low priority SNs that generate less important data at lower rates.

To control a SNs priority, the effect is exploited that a receiver ignores date packets with lower signal strength if at the same time a packet with larger signal strength is being transmitted. Hence, SNs are given high priority by placing them in the vicinity of a CH, whereas a SN is given a lower priority by placing it farther apart from CHs.

As mentioned, our objective is to maximize the total rate of received date packets. If the network is lightly loaded, a CH will receive most of the packets. When the load is high, CHs should receive packets only from near-by, high priority SNs, and ignore the distant ones. Note that packets from distance SNs would be lost anyway due to a strong interference caused by close high priority SNs.

We illustrate our design objective with an example of a surveillance system, originally defined in [91], see the scenario shown in Figure A.1. An underground car park is filled with SNs of different type. Some are of low priority, like those for temperature and humidity measurements; they generate very low traffic (< 1 kbit/s), the loss of one or a few of their data packets is acceptable. Other SNs are of high priority, like seismic, infrared and microphone SNs that are used to detect movements of an intruder (≈ 10 kbit/s traffic), and cameras that are transmitting live videos from the area ($\approx 100 - 400$ kbit/s traffic).

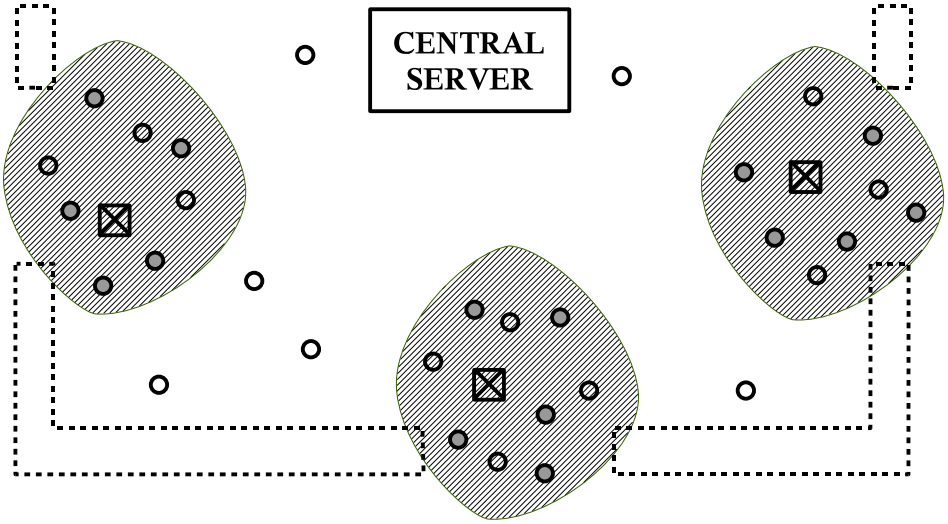


Figure A.1: An illustration of a transmit-only sensor network in an intrusion detection scenario. SNs are denoted with circles, and CHs with crossed boxes. Empty circles are low-traffic and low-priority SNs and solid circles are high-traffic and high-priority (e.g. video) SNs. For the given CH positions, the shaded areas indicate potential locations for high priority SNs.

High-priority SNs, like cameras and movement detection SNs are placed near the CHs. When

the traffic is high, there will be many collisions between data packets. As SNs are unaware of the traffic intensity collisions cannot be prevented. Nevertheless, if high-priority SNs are close to the CHs, interference from distant SNs are going to be low compared to the received signal power such that most of their packets will be received. On the contrary, low-priority SNs may be placed significantly farther away from the CHs; their packets will be correctly received only when no intrusion is detected (in this case the total traffic is low), which is assumed to be sufficient for the application.

A similar application is a fire detection sensor network [79]: high-priority SNs detect fire and monitor its spreading; various low-priority SNs monitor temperature, humidity, solar power, etc.

A.3 Detailed Description of The Proposed Architectures

A.3.1 Simple CHR Architecture

A CHR consisting of a simple noncoherent receiver in the sense of [95] is called a *simple CHR*, it can receive only one packet at a time. While a CH is receiving a packet from a distant SN, a parallel transmission may start from a near-by, high-priority SN, since SNs are not aware of an ongoing transmissions. This new transmission will interfere and may corrupt the packet being received. At the same time, the CH will not be able to receive the interfering, high-priority packet since the CH's receiving circuit was busy when its transmission started. Hence, both low- and high-priority packets are lost. We next propose two simple CHR architectures that address this issue and improve the performance.

A.3.2 Adaptive-Threshold CH Architecture

We first propose the adaptive CHR architecture. If the signal strength of the received packet is lower than the detection threshold E_d, then the packet will be ignored. Otherwise, the CHR will try to receive it. We say that a packet is **detected** if the received signal strength is larger than E_d. The concept is illustrated in Figure A.2. Packets from SNs that are outside of the shaded area have the received signal strength lower than E_d, hence they will be dropped. Only packets from SNs inside the circle will be detected. The main reason why a CH should not start receiving a low signal-strength packet is, as explained in Section A.3.1, that in case of a collision, both packets will be lost.

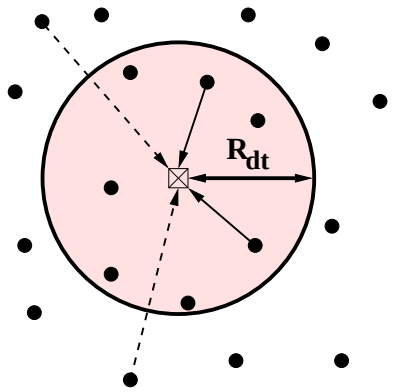


Figure A.2: Illustration of the detection threshold principle: Transmissions of SNs whose received signal strength at the CH is higher than E_d are indicated with solid lines. Transmissions of SNs whose received signal strength is below E_d are indicated with dashed lines. Only packets whose sources are in the shaded circle will be detected.

Effect of Detection Threshold

To better understand the adaptive CHR, we define a simple model of the sensor network and its components and analyze the total data throughput. First we assume that the traffic generated by the SNs is Poisson distributed. This assumption is reasonable because the optimal medium access scheme for the SNs is to defer each transmission for a random time [74]. In addition, we assume a simplified physical layer model: if a CH receives a packet from a SN with signal strength E_r, and if an interfering packet has the signal the strength $E_r - \Delta$, then the received packet will be lost. This simplification neglects the impact of multiple interferers, but as we shall see, it provides results that fit well with simulation.

We consider again the scenario depicted in Figure A.2, with one CH and n randomly distributed SNs. Packets from the i-th SN, SN_i, have a signal strength $E_{r,i}$ at the CH. The SNs generate data packets at independent random time instances that are characterized by a Poisson distribution with the rate parameter λ_i. The value $\lambda = 1$ corresponds to zero delay between successive packets, i.e., to the maximum sustainable rate supported by a SN's transmitter. Hence, the total traffic load normalized to the sustainable rate, which is generated by the SNs inside the circle is

$$\lambda_a(E_d) = \sum_{i:E_{r,i} \geq E_d} \lambda_i. \tag{A.1}$$

We first estimate the probability that the receiver is idle at any given moment in time. The receiver is idle if it has finished decoding a previous packet (successfully or unsuccessfully), and if no other packet has arrived in the meantime with a received signal strength larger than the

detection threshold, E_d. The state of the receiver (busy or idle) is a stationary process in time that can be described by a continuous Markov chain [11], its stationary idle probability is given by

$$\overline{P}_{\text{idle}}(E_d) = \frac{1}{1 + \lambda_a(E_d)}. \tag{A.2}$$

Next we calculate the probability that a packet transmitted by SN_i will successfully be received. This happens if the receiver is idle at the time SN_i starts to transmit, and if the data packet does not overlap with any packet whose received signal strength is larger than $E_{r,i} - \Delta$. We assume that all packets have a fixed duration T_p which is is normalized to one, i.e, $T_p = 1$. Similar to non-slotted ALOHA, if a packet is sent at time zero, then any other packet arriving within the interval $[-1, 1]$ will interfere and destroy the packet. Therefore, the probability to capture the packet of the i-th SN is

$$P_{c,i}(E_d) = \overline{P}_{\text{idle}}(E_d)\ \exp\left(-2 \sum_{j:E_{r,j} \geq E_{r,i} - \Delta} \lambda_j\right). \tag{A.3}$$

With this, the rate of packets that are received from SN_i is $\lambda_i * P_{c,i}(E_d)$ and the aggregate throughput of all SNs normalized to the sustainable transmitter rate is

$$X(E_d) = \sum_{i:P_{r,i} \geq E_d} \lambda_i\ P_{c,i}(E_d) \tag{A.4}$$

The problem of maximizing $X(E_d)$ can be solved numerically. We solved it for a large number of topologies and traffic rates and we found that the optimal value for $X(E_d)$ is 0.25 and is obtained if the load λ_a is maintained at a level of 0.75. In other words, to maximize the normalized aggregate throughput X we should estimate $\lambda_a(E_d)$ and vary E_d to obtain $\lambda_a = 0.75$.

The above analytical model shall be verified by means of simulation. We randomly distribute 50 SNs in a 40×40 m square and measure the aggregate throughput of the system for different loads. We see in Figure A.3 that the throughput is maximal when the normalized aggregate traffic load is around 0.75. At that point, the normalized aggregate throughput of the network, for simplicity called utilization, is slightly above 0.25.

Traffic Load Estimation

Now a simple method is described for tracking λ_a and X and to adapt E_d in order to keep the utilization, X, at a maximum. We denote with $\hat{T}_{l,k}$ the estimated average idle time of a receiver,

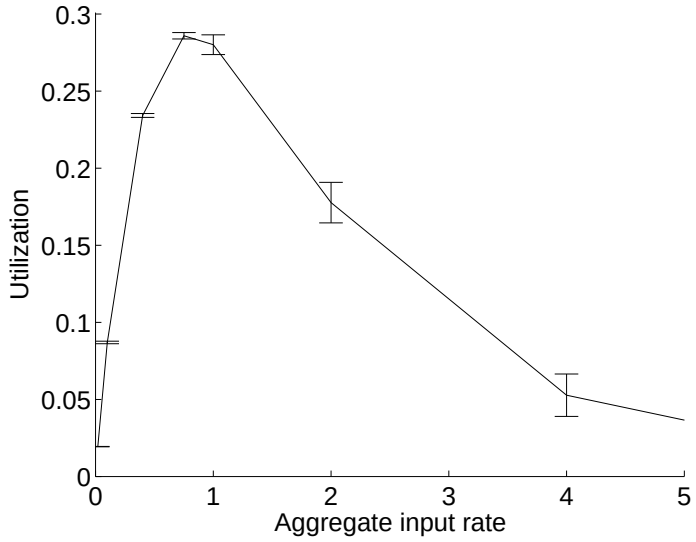


Figure A.3: Simulation results for a network with 50 SNs uniformly distributed in a 40×40 m square and one CH in the middle. We plot the aggregate throughput as a function of the aggregate SN traffic load, both normalized by the data rate per SN.

i.e. the average time between two packets that are detected, see Figure A.4; the index k denotes the iteration step. For this purpose, we keep track of the time instant T_{det} when the last packet has been detected and of the time instant T_{end} when the last detected packet ended (successfully or unsuccessfully received). With this, the average idle time can be computed iteratively as

$$T_{l,k} = \alpha T_{l,k-1} + (1-\alpha)(T_{\text{det}} - T_{\text{end}}). \tag{A.5}$$

The estimate $\hat{\lambda}_a$ of the load, λ_a, at the k-th iteration can be computed as the ratio of the packet duration, T_p, to the packet duration $T_{l,k}$, i.e., $\hat{\lambda}_a = T_p/T_{l,k}$.

A similar estimation can be computed for the utilization X. Therefor, we denote with $\hat{T}_{u,k}$ the estimated average time between two successful receptions at the k-th iteration and the time instant at the end of the last successful packet transmission is as $T_{\text{end\_succ}}$. At the end of the next successful packet transmission $T_{\text{next\_succ}}$, we update

$$T_{u,k} = \alpha T_{u,k-1} + (1-\alpha)(T_{\text{next\_succ}} - T_{\text{end\_succ}}). \tag{A.6}$$

The estimate $\hat{X}$ of the utilization X is the ratio of the packet duration T_p to $T_{u,k}$, i.e., $\hat{X} = T_p/T_{u,k}$. The weighting parameter α was set to $\alpha = 0.95$ for both, determination of T_u and T_l.

A CH needs also to learn about existing SNs and their received signal strengths. This is done

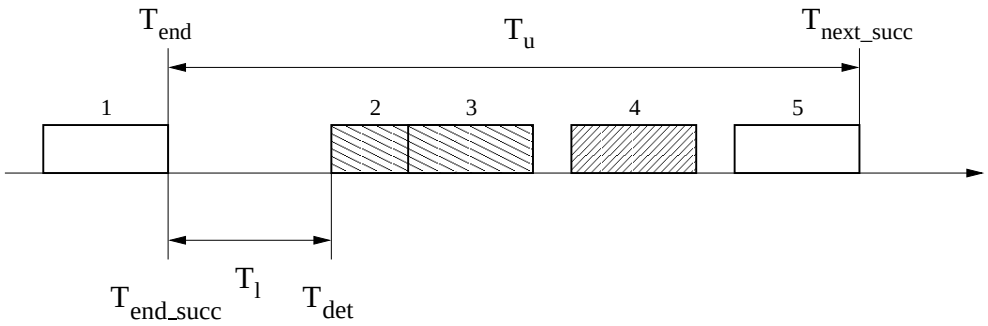


Figure A.4: An example of packet arrivals that illustrates the adaptation mechanism. Packet 1 has arrived and is well received. Packet 2 was being received when it collided with packet 3 and is discarded. Packet 4 is not detected because it arrives from a node outside of the detection circle. Finally, packet 5 again is well received.

during packet receptions. Note that a CH does not necessarily need to successfully receive a whole packet to perform this task. It might be sufficient to decode the header and to estimate the signal strength. As a result of this estimation a CH keeps a list of active SNs and their received signal strengths $\{i, E_{r,i}\}_{i=1,\cdots,n}$, ordered with respect to signal strengths.

The key goal of the adaptive threshold algorithm is to keep the utilization at 0.25 and the detected load at 0.75. It should be sufficient to use the estimated load, $\hat{\lambda}_a$, but we use both estimated load and estimated utilization, $\hat{X}$, to have better robustness against estimation errors. The adaptive threshold algorithm is initialized with the value $E_d = 0$, thus it is able to detect any transmission. The variables T_l and T_u are updated each time a packet is received and the list of SNs and their received signal strengths. If the estimated load $\hat{\lambda}_a$ exceeds the level 0.75 and if at the same time the utilization estimate $\hat{X}$c drops below 0.25 in can be concluded that the actual operating point is at the right side of the top of the curve in Figure A.4, i.e., that the detection threshold, E_d, is too small and has to be increased in order to decrease the traffic load λ_a.

The detection threshold is updated every time a new packet is sensed. It is important to notice that it is more critical to chose to large a value for the detection threshold than using to small a value; this is due to the shape of the curve in Figure A.3. If we overestimate the detection threshold when the total traffic is low (at the left from the curves peak), then we would further decrease the traffic load, and hence further decrease the utilization. Similarly, when the traffic load is high (at the right from the curves peak) and a value is chosen for the detection threshold below the optimum value, then the total number of detected packets is decreased and with this the utilization X.

For this reason, we propose a conservative decrease of E_d: only if it happens for four packets in a row that the detected load λ_a is larger than 0.75 and the utilization lower than 0.25, then the

detection threshold should be increased such that one SN is removed from the SN list; in other words, if $E_d = E_{r,i}$ and $i > 1$, then we update $E_d = E_{r,i-1}$.

A less conservative procedure is chosen to decre the detection threshold. The first time when the detected load is lower than 0.75, and the utilization is lower than 0.25, we set $E_d = E_{r,i+1}$ (if $i < n$, or else $E_d = 0$).

Another important point is to keep updating T_l and T_u even when no packet arrivals are detected. If the traffic intensity suddenly drops, or nearby SNs cease transmitting, to large a detection threshold can result. A method to avoid this, is to gradually decrease the detection threshold while no packets are detected. This is achieved by setting

$$\hat{\lambda}_a = \frac{T_p}{\alpha T_l + (1-\alpha)(T_{\text{det}} - T_{\text{end}})}, \tag{A.7}$$

$$\hat{X} = \frac{T_p}{\alpha T_u + (1-\alpha)(T_{\text{det}} - T_{\text{end\_succ}})}, \tag{A.8}$$

where T_{det} is the beginning time of a detected packet and $T_{\text{end\_succ}}$ is the end of the last packet that was successfully received, see Figure A.4.

A.3.3 Switched CH Architecture

As explained in the previous section, the main reason for low efficiency of a CHR is that if it starts receiving a packet with small signal strength and a packet with a stronger signal arrives during this reception, then the weaker packet will be dropped due to interference and the stronger one will not be received because the receiving circuit was busy with the weaker packet.

The most general way to solve this problem is to include several receiving circuits in parallel so that a CH can cope with all arriving packets. This is too expensive and not necessary in most of the cases. We propose a less complex receiver which consists of a simple receiver circuit and an additional acquisition circuit. This additional circuit constantly monitors the wireless medium for a newly arriving packet. If a transmission of a new packet starts and if its signal is stronger than the signal of the packet currently being received, then the CH stops receiving the current packet and switches to the new, stronger one. We call this architecture *switched* CHR *architecture*.

A.4 Performance Comparison

To analyze the performance of the proposed architectures we have implemented a cross-layer, discrete-event simulator that models the UWB physical layer, the multi-access schemes, and the proposed CH architectures with the greatest possible detail. A description of the simulator can be found in [72].

At the physical layer, pulses are sent with an energy level of 30 pJ. As described in [95], this implies a link data rate of 2.5 Mbps and a range of approximately 30 m (using LOS propagation model from [28]). Thus all SNs can be heard by the CHs.

We evaluate the performance of two networks, each having 50 SNs distributed uniformly in a 40×40 m square. The first network has one CH in the center of the square, and the other two CHs located at (10 m,20 m) and (30 m, 20 m). All SNs generate Poisson traffic with the same intensity.

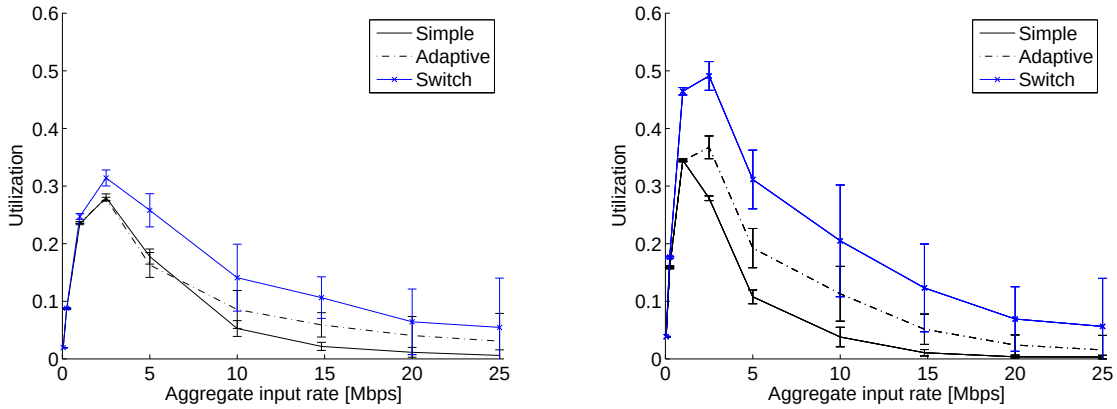


Figure A.5: Aggregate utilization X is plotted as function of the total load. The underlying scenario consists of 50 SNs uniformly distributed in a 40×40 m square. All SNs generate Poisson traffic with the same intensity. The left figure corresponds to one CH located at the center of the square. The right figure corresponds to two CHs located at (10 m, 20 m) and (30 m, 20 m).

We vary the packet generation rate and measure the number of captured packets. The results are depicted in Figure A.5, where the performance of the two proposed architectures is compared to the one of a the simple CHR, defined in Section A.3.1. When the aggregate traffic rate is small, there is no difference between the three architectures, since there is almost no interference between the SNs and it is possible to detect and receive packets from all SNs. However, as the input rate increases above 10 Mbps, the performance differences become significant, and the total number of packets received by the adaptive CHR is twice the one of the simple CHR. We

can also see that in case of the switched CHR we can additionally double the performance, note that this improvement is achieved at the cost of a slightly more expensive hardware.

A.5 Conclusions

We described a system architecture for a low-cost UWB sensor network with transmit-only sensor nodes. As most of the design flexibility of the network is in the CHs, we proposed two novel CHR architectures that maximize the number of received packets: a so called adaptive CHR and a switched CHR. The adaptive CHR architecture requires the same hardware as the simple non-adaptive architecture, therefore has almost the same implementation cost. The switched CHR architecture introduces an additional detection and synchronization circuit; it is thus expected to be more expensive.

We studied the performance of the proposed CHR architectures by means of simulation and found that for low traffic load the proposed architectures perform as well as the simple CHR. For high traffic load, the switched CHR is twice better than the adaptive CHR, which in turn is twice better than the simple CHR.

The CHR design presented in this paper is created for networks with transmit-only SNs. The presented numerical results are obtained for a physical layer which is based on pulsed UWB radio. Nevertheless, the ideas are applicable to other physical layers as well, where it is prohibitively expensive to put a receiver in each sensor node.

Index

achievable rate or power allocation, 137
attenuation, 57, 61, 94

background noise, 57, 61

capacity, *see* sum of rates
capacity metric, 65
Chip size, T_c, 93

efficiency index, 64

fairness metric, 65
Frame size, T_f, 93
free-disposal property, 42

leximin larger, 32
leximin maximal, 32
linear rate function, 94
long-term average power, 60, 62

MAC protocol, 58–59, 96–97
max-min achievable, 33
max-min fairness, 31, 56, 64
max-min fairness index, 65
max-min programming, 39
min-max fairness, 31
minimum potential delay, 56
minimum average rates, 136
mobility, 98
model, 56–63
α, schedule, 58–59, 61
β, orthogonality factor, 95
C_r, routing overhead, 97
η, background noise, 57, 61
F, flow matrix, 59, 62, 97
$\mathcal{F}$, set of feasible average rates, 62
$\mathbf{f}$, instantaneous end-to-end flow rates, 61
$\bar{\mathbf{f}}$, average end-to-end flow rates, 62
$\overline{\mathbf{F}}^{MIN}$, minimum average rates, 136
$\mathbf{f}^{\mathrm{nm}}$, average flow rate without mobility overhead, 98
$\mathbf{f}^{\mathrm{nr}}$, average flow rate without routing overhead, 97
$\mathbf{f}^{\mathrm{nrm}}$, average flow rate without routing and mobility overhead, 97, 98
h, attenuation, 57, 61, 94
mobility, 99
r_i, number of hops, 97
$\mathbf{P}^{MAX}$, transmission power constraints, 62
$\overline{\mathbf{P}}^{MAX}$, long-term average power constraints, 62
$\mathbf{p}$, transmission power, 60, 61, 96
$\bar{\mathbf{p}}$, long-term average power, 60, 62
q, probability of packet loss due to mobility, 98
R, routing matrix, 59, 62, 97
$r()$, rate function, 57, 61, 94
S, system state, 57, 61
$\mathbf{SINR}$, signal-to-interference-and-noise ra-

tio, 61
$\mathcal{T}$, set of feasible average transport rates, 62
$\mathbf{x}$, instantaneous rate, 61
$\bar{\mathbf{x}}$, average rate, 61
$\mathbf{y}$, rate on paths, 61

optimal routing, power allocation and scheduling, 100
orthogonality factor, 95

Pareto efficient, 135, 137
peak power, 60
power constraints, 60, 95
proportional fairness, 36, 56
pulse position modulation, 93

rate criteria, 63, 64
rate function, 57, 61, 94
rates maximization, *see* sum of rates
routing, 59, 62, 97–98
DIR - direct routing, 97
MELR - minimum energy loss routing, 97, 101
MER - minimum energy routing, 101
routing overhead, 97

schedule, 58–59, 61, 96–97
set of feasible average consumed powers, 137
set of feasible rates, 137
solidarity property, 66
sum of rates, 56, 64
system state, 57, 61

transmission power, 60, 61, 96
transport capacity, 64
transport capacity metric, 65
transport max-min fairness index, 65
ξ-transport utility metric, 66
transport-max-min fairness, 64
transport-utility fairness, 64

ξ-utility, 65
utility fairness, 36
utility fairness, 56, 64

water-filling, 42
Win-Scholtz physical model, 93

Appendix B

Curriculum Vitae and List of Publications

Božidar Radunović was born in Belgrade, Yugoslavia. He received the BSc degree in electrical engineering in 1999 from the School of Electrical Engineering, Belgrade, Yugoslavia.

From 1999-2000 he attended the Doctoral School in Communication Systems at EPFL, Switzerland.

In 2000 he joined the Laboratory of Communications and Applications (LCA) at the School of Computer and Communication Sciences, EPFL. From 2000-2005 he participated in Swiss National Science Foundation research project called Mobile Information & Communication Systems, where he contributed in the field of cross-layer network design. His current research include design of wireless multi-hop networks, performance evaluation, efficiency and fairness in wireless networks and UWB (ultra-wide band) wireless networks.

PUBLICATIONS

- Radunovic, B. and Le Boudec, J.-Y., "Rate Performance Objectives of Multihop Wireless Networks" *In IEEE Transactions on Mobile Computing*, October 2004, Vol. 3 No. 4, 334-349.

- Radunovic, B. and Le Boudec, J.-Y., "Optimal Power Control, Scheduling and Routing in UWB Networks" *In IEEE Journal on Selected Areas in Communications*, September 2004, Vol. 22 No. 7.

- Radunovic, B. and Le Boudec, J.-Y., "A Unified Framework for Max-Min and Min-Max Fairness with Applications" *In IEEE Transactions on Networking*, Submitted.

- Merz, R. and Widmer, J. and Le Boudec, J.-Y. and Radunovic, B. "A Joint PHY/MAC Architecture for Low-Radiated Power TH-UWB Wireless Ad-Hoc Networks", *In Wireless Communications and Mobile Computing Journal, Special Issue on Ultrawideband (UWB) Communications*, June 2005.

- Milutinovic, V. and Knezevic, P. and Radunovic B., Casselman, S. and Schewel, J. "Obelix Searches Internet Using Customer Data," *IEEE COMPUTER*, July 2000.

- Radunović, B. "Convergence in Hausdorff Metric Preserves Geometric Shape," *Matematički vesnik 48(1996)*, pp. 25-28

- Radunovic, B. and Le Boudec, J.-Y., "Rate Performance Objectives of Multi-hop Wireless Networks, " *Proceedings of INFOCOM 04*, Hong Kong, March 2004.

- Radunovic, B. and Le Boudec, J.-Y. "Power Control is Not Required for Wireless Networks in the Linear Regime" *Proceedings of WoWMoM*, Taormina, Italy, June 2005.

- Radunovic, B. and Le Boudec, J.-Y. "Rate Adaptation, Power Adaptation and Mutual Exclusion in Variable Rate Wireless Networks" *Proceedings of RAWNET*, Riva Del Garda, Italy, April 2005.

- Merz R., Le Boudec J.-Y., Widmer J, Radunović B., "A Rate-Adaptive MAC Protocol for Low-Power Ultra-Wide Band Ad-hoc Networks, " *Proceedings of Ad-Hoc Now 04, 3rd International Conference on AD-HOC Networks* , Vancouver, British Columbia, Canada, July 2004.

- Le Boudec J.-Y., Merz R., Radunović B., Widmer J, "DCC-MAC: A Decentralized MAC Protocol for 802.15.4a-like UWB Mobile Ad-Hoc Networks Based on Dynamic Channel Coding", *Proceedings of Broadnets*, San Jose, October 2004.

- Radunovic, B. and Le Boudec, J.-Y., "Joint Scheduling, Power Control and Routing in Symmetric, One-dimensional, Multi-hop Wireless Networks, " *Proceedings of WiOpt Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks*, INRIA Sophia-Antipolis, France, March 2003.

- Radunovic, B. and Le Boudec, J.-Y. "A Unified Framework for Max-Min and Min-Max Fairness with Applications, " *Proceedings of 40th Annual Allerton Conference on Communication, Control, and Computing*, Allerton, IL, October 2002.

- Kencl, L. and Radunovic, B. “A General Method for Finding the Most Economical Distributed Router Architecture,” *Proceedings of the CNDS’02*, San Antonio, Texas, USA. January 27-31, 2002.

- Knezevic, P. and Radunovic, B. and Nikolic, N. and Jovanovic, T. and Milanov, D., Nikolic, M. and Milutinovic, V. and Casselman, S. and Schewel, J. “The Architecture of the Obelix - An Improved Internet Search Engine,” *Proceedings of the HICSS-33*, Maui, Hawaii, USA. January 4-7, 2000.

- Knezevic, P. and Radunovic, B. and Elhouni, A. “Issues in FPGA-based Configurable Computer Design for Operating Systems and Multimedia Applications”, *Proceedings of the HICSS-32*, Maui, Hawaii, USA. 1999.

- Radunovic, B. “An Overview of Advances in Reconfigurable Computing Systems”, *HICSS-32*, Maui, Hawaii, USA. 1999.

- Radunovic, B. and Milutinovic, V. “A Survey of Reconfigurable Computing Architectures”, FPL ’98, Tallinn, Estonia, 1998, 376-385

- B. Radunovic, L. Truong, and M. Weisenhorn, “On architectures of transmit-only, UWB-based wireless sensor networks,” IBM Research Report RZ 3596, 2005.

- B. Radunovic, “A cross-layer system simulator for UWBbased wireless sensor networks,” IBM Research Report RZ 3594, Feb. 2005.